BMJ Open

Prognostic risk factor analysis and construction of a nomogram for elderly primary liver cancer patients based on SEER database

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-051946
Article Type:	Original research
Date Submitted by the Author:	01-Apr-2021
Complete List of Authors:	Li, Fangyuan; The first People's Hospital of Yuhang District, Department of Medical Oncology Zheng, Ting; The first People's Hospital of Yuhang District, Department of Medical Oncology Gu, Xuewei; Zhuji People's Hospital of Zhejiang Province (Zhuji Affiliated Hospital of Shaoxing University), Department of Gastroenterology
Keywords:	Cancer genetics < GENETICS, Adult oncology < ONCOLOGY, Gene therapy < ONCOLOGY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Prognostic risk factor analysis and construction of a nomogram for elderly primary liver cancer patients based on SEER database

Fangyuan Li[1], Ting Zheng[1], Xuewei Gu[2,*]

[1] Department of Medical Oncology, The first People's Hospital of Yuhang District, Hangzhou, 369 Yingbin Road,Yuhang District, Hangzhou, 311100, Zhejiang, PR China.

[2] Department of Gastroenterology, Zhuji People's Hospital of Zhejiang Province (Zhuji Affiliated Hospital of Shaoxing University), 9 Jianmin Road, Taozhu street, Zhuji, 311800, Zhejiang, China.

*Corresponding Author（Xuewei Gu）: Department of Gastroenterology, Zhuji People's Hospital of Zhejiang Province (Zhuji Affiliated Hospital of Shaoxing University), 9 Jianmin Road, Taozhu street, Zhuji, 311800, Zhejiang, China.

E-mail: gxwgxw1234567@163.com

Running title: Nomogram model for primary liver cancer in elderly patients

Key words: nomogram, primary liver cancer, elderly, database

Abstract

Objective: To evaluate the risk factors and construct a nomogram model for the prognosis of primary liver cancer in the elderly based on the data from the US SEER database. **Methods:** The latest data of primary liver cancer patients were extracted from the SEER database with SEER*STAT software, and the required variables were included. The data was screened and then divided into training cohort and validation cohort. Variates were first screened by univariate and multivariate COX analysis, and variates with then statistical significance were included the construction of a nomogram model. The C-Index, ROC and calibration curves were used to evaluate the model. **Results:** A total of 10824 eligible cases from 2004 to 2017 were extracted, among which, 7757 cases were included in the training cohort and 3247 in the validation cohort. The C-Index of the model was 0.747 (training cohort) and 0.773 (validation cohort). The 3-year AUCs of the training and validation cohort were 0.760 and 0.750, while the 5-year AUCs were 0.761 and 0.748. The calibration curves showed an ideal calibration of the constructed model. **Conclusions:** The nomogram model constructed followed by COX regression analysis showed ideal calibration and discrimination property, and can provide a reference for clinical application of elderly primary liver cancer in the future.

Strength and limitations of this study:

Strength:

1.Collected a large and sufficient number of cases from SEER database of elderly liver cancer.

2.Constructed a novel and ideal prognostic model for elderly liver cancer

Limitations:

1.All cases were collected form one database and selection bias might exist.

2.Some classification carried out by SEER database was not very specific.

3.Information such as ancillary tests was absent from the SEER database.

Keywords: nomogram, primary liver cancer, elderly, database

Introduction

Primary liver cancer is currently the sixth most common cancer worldwide and, according to epidemiological surveys, is the fourth leading cause of cancer-related deaths globally, representing a major threat to the health of the entire human population [1,2]. Furthermore, many studies have pointed out that although middle-aged (30-59 years old) or young (< 30 years old) patients with primary liver cancer are not uncommon worldwide, the average age of diagnosis of the disease is 60 years old and, in contrast to the yearly decrease in the age-standardized incidence rate (ASR) of young patients, the incidence of elderly patients has continuously increased in more than half of

the countries and regions during the last 30 years [3-5]. The global population expansion, increasing aging, as well as obesity, diabetes, overmedication and lagging effects of HBV infection in the elderly may be responsible for the high or even increasing ASR of primary liver cancer in elderly patients, creating a heavy burden on the health of all countries [6-8]. In terms of treatment, surgery remains the first choice for primary liver cancer. Therefore, based on the epidemiological characteristics and treatment modalities of the disease, it is necessary to conduct an accurate prognostic assessment of elderly patients with primary liver cancer to guide clinical work. However, the different pathological types and heterogeneity of the disease make prognostic assessment still difficult.

Recently, the nomogram model has shown superior predictive performance over the traditional TNM staging because of its convenient modeling method and the ability to incorporate multiple variables, thus gaining widespread popularity [9,10]. This study intended to use a nomogram model to analyze the risk factors affecting elderly primary liver cancer in the SEER database and to predict the prognosis of the disease. The assessment performance of the model was analyzed by the test of discrimination and calibration to establish an optimal assessment system for clinical work such as the treatment of elderly primary liver cancer.

Methods and data

1. Ethical statement

Informed consent was not required from patients to obtain data from the US SEER database since cancer is publicly reportable in every state in the United States.

2. Case selection

Case data of primary liver cancer with complete follow-up records were selected from the 2004-2017 SEER database (SEER research data, 18 Registries, Nov 2019 Sub (2000-2017)) using SEER*Stat 8.3.6. The extracted variables included race, year of diagnosis, age, sex, primary site, histologic type, grade, TNM stage, tumor size, surgery on the primary site, survival time, cause of

death and survival status. Among them, patients who were older than 65 years old were selected; Asian and Pacific Islanders, American Indians and Alaskan natives were included in Asian and others for the race variable; liver or intrahepatic bile duct (IBD) was selected as the primary site; intrahepatic cholangiocarcinoma (ICC), hepatocellular carcinoma (HCC), and combined hepatic carcinoma (CHC) were selected as the histologic type; and local destruction included photodynamic therapy (PDT), percutaneous ethanol injection (PEI) and radiofrequency ablation (RFA), etc.

3. Statistical processing

The optimal cut-off value of tumor size was analyzed using the X-Tile software, and the variable was classified into < 46 mm, 46-81 mm and > 81 mm according to high, medium and low risk. After that, the total cases were randomly assigned to a training or a validation cohort at a ratio of 7:3 using SPSS 18.0 by random number 20200222, followed by the collection of baseline information. Univariate and multivariate (Forward: LR) COX analyses were performed using the R software or SPSS to screen for statistically significant variables to construct a nomogram, based on which C-Index, ROC curves and the area under the curve (AUC) were figured out. Calibration curves of the model for 3 and 5 years were plotted with the R software after Bootstrap sampling for 1000 times. $P < 0.05$ was considered statistically significant.

Results

1. Clinical characteristics of the cases

A total of 10,824 elderly cases with primary liver cancer were extracted in accordance with screening conditions, with 7,757 included in the training cohort and 3,247 in the validation one. Among them, the majority of patients were male (67.5%), white (71.3%), with primary site in the liver (87.8%), HCC (84.7%), grade II (46.6%), T1 (46.5%), N0 (91.6%), M0 (88.7%) and unoperated (56.5%) (Table 1).

2. Screening for prognostic risk factors

Univariate COX regression analysis of the training cohort proved the variates of age, sex, race, histological type, grade, TNM stage, surgery and tumor size to be statistically significant ($P < 0.05$), which were included in the follow-up multivariate COX analysis. However, the primary site was excluded according to the analysis ($P = 0.232$) (Table 2). Subsequently, the sex variable was further excluded from the experiment by Forward: LR multivariate COX (Table 3). In conclusion, age, race, histological type, grade, TNM stage, surgery, and tumor size were independent risk factors that affect the prognosis of elderly patients from the extracted data with primary liver cancer, which could be used to construct a subsequent nomogram prediction model.

4. Nomogram model construction and verification

The independent risk factors affecting prognosis derived from the above analysis were incorporated to construct a 3- and 5-year nomogram prediction model for elderly primary liver cancer, and the total score was calculated by aggregating the scores of each variable to predict the survival rate of patients at 3 and 5 years (Figure 1). It can be seen that the surgery on the primary site was the most important factor affecting the score in this model, followed by tumor size, TNM stage and age. The AUC was calculated after plotting the ROC curves of the training and validation cohorts. Specifically, the AUC is 0.760 (3 years) and 0.761 (5 years) in the training cohort and 0.750 (3 years) and 0.748 (5 years) in the validation cohort (Figure 2). Furthermore, the model showed an ideal calibration for 3 and 5 years in both groups after creating the calibration curves for the training and validation cohorts (Figure 3).

Discussion

Analysis of cases revealed that male patients accounted for more than 60% of elderly patients

with primary liver cancer. Some statistics have presented that the mean annual change rate of men suffering from the disease is higher than that of women (3.7% vs. 2.7%) in the United States [11]. In China, a population-based study of hepatic carcinoma in Zhejiang Province demonstrated that the ASR for hepatic carcinoma was 33.24 in men compared to 1.21 in women [12]. Not only differences in lifestyle---including alcohol consumption and smoking---have led to higher cancer rates in men, but different physiological conditions such as hormone secretion and even genetic differences may be responsible for these epidemiological differences [13]. Therefore, it has been put forward that gender is a critical biological variable that should be considered in all studies aimed at improving carcinoma [14]. Analysis of baseline data also suggested that the population of elderly primary liver cancer was predominantly white and mostly with the primary site in the liver, HCC histological type, grade II (moderately differentiated), T1 and without lymph node metastasis or distant metastasis. Moreover, in this population, more than half of the cases were not treated surgically. The reason for this phenomenon may be that most of the patients were over 60 years old at the time of diagnosis, missing the best timing for radical surgery, together with the decline in physical function as well as intolerance to surgery led to a palliative treatment for most patients.

Based on further univariate and multivariate COX analyses, several independent risk factors that affect the prognosis of the disease were sifted out, including age, race, histological type, grade, TNM stage, surgery and tumor size. Sex, though not negligible as previously mentioned, was found not to be the main factor affecting prognosis in this population after comprehensive analysis, which is consistent with several current retrospective studies on hepatic carcinoma [15-17]. In terms of histological type, it is evident that CHC has a worse prognosis than the common HCC, which shows a lower incidence but a higher malignancy [18-19]. Analysis of the age factor revealed that the higher the age group of the patient, the worse the prognosis, suggesting a linear negative correlation trend. The nomogram model also indicated that the surgery option was the most crucial factor influencing the prognosis of the disease. The patients who underwent liver transplantation, though small in number, showed a relatively good prognosis, followed by resection or lobectomy and then by local destruction. In contrast, patients without surgery showed a poor prognosis. This factor alone reduced the 3- and 5-year predicted

survival rates to less than 50%, suggesting that the invention of new methods or enhanced surgery is still an urgent issue for improving the prognosis of elderly primary liver cancer. The influence of other factors on prognosis is basically in line with the current consensus that the worse the grade, the higher the T-stage, the occurrence of lymph node metastasis, the occurrence of distant metastasis and the larger the tumor, the worse the prognosis of the patient.

After that, the performance of the established model was evaluated by C-Index, ROC curves and calibration curves. A nomogram model is considered to have good discrimination if its C-Index and AUC exceed 0.7 [20,21]. As the model constructed in this study had these two indicators above 0.7 in both the training and validation cohorts and the calibration plots scatter in accordance with the reference line, it could be considered that the model has good discrimination and calibration and hence the capacity to predict the prognosis of the disease.

However, this study also has shortcomings. First, the cases were all from the US SEER database, which is not representative of regions other than the United States and is subject to selection bias. In addition, the case data included in this database lacked some important ancillary tests related to the diagnosis and treatment of liver cancer, such as CEA, AST and vascular invasion.

In conclusion, a nomogram model with favorable prediction was developed by using the case data from the SEER database after performing univariate and multivariate COX screening, which could provide a reference for the future diagnosis and treatment of elderly patients with primary liver cancer.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Conflicts of interest

The authors declare there are no competing interests.

Author contributions

Fangyuan Li wrote and revised the manuscript; Ting Zheng conducted most of the analysis of data; Xuewei Gu reviewed the manuscript; All authors read and approved the final manuscript.

Availability of data and material

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Patient and Public Involvement

No patient involved.

Ethics approval and consent to participate

Informed consent was not required from patients to obtain data from the US SEER database since cancer is publicly reportable in every state in the United States.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;70(1):7-30. doi:10.3322/caac.21590

2. Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. 68, 394–424. https ://doi. org/10.3322/caac.21492 (2018).

3. Liu Z, Suo C, Mao X, et al. Global incidence trends in primary liver cancer by age at diagnosis, sex, region, and etiology, 1990-2017. Cancer. 2020;126(10):2267-2278. doi:10.1002 /cncr.32789

4. Kanwal F, Kramer JR, Mapakshi S, et al. Risk of hepatocellular cancer in patients with non-alcoholic fatty liver disease. Gastroenterology. 2018;155:1828-1837.e2.

5. Zhang S, Zheng R, Zeng H, Chen W. The incidence differences among sex, geographical areas and mean age of diagnosis for liver cancer in China, 1989-2008 [in Chinese]. Zhonghua Yu Fang Yi Xue Za Zhi. 2014;48:355-360.

6. Baumeister SE, Schlesinger S, Aleksandrova K, et al. Association between physical activity and risk of hepatobiliary cancers: a multinational cohort study. J Hepatol. 2019;70:885-892.

7. Klil-Drori AJ, Azoulay L, Pollak MN. Cancer, obesity, diabetes, and antidiabetic drugs: is the fog clearing? Nat Rev Clin Oncol. 2017;14:85-99.

8. Smittenaar CR, Petersen KA, Stewart K, Moitt N. Cancer incidence and mortality projections in the UK until 2035. Br J Cancer. 2016;115(9):1147-1155. doi:10.1038/bjc.2016.304

9. Chen, S., et al., Development and Validation of a Nomogram for Predicting Survival in Male Patients With Breast Cancer. Front Oncol, 2019. 9: p. 361

10. Zhang Y, Hong YK, Zhuang DW, He XJ, Lin ME. Bladder cancer survival nomogram: Development and validation of a prediction tool, using the SEER and TCGA databases. Medicine (Baltimore). 2019 Nov;98(44):e17725.

11. Massarweh NN, El-Serag HB. Epidemiology of Hepatocellular Carcinoma and Intrahepatic Cholangiocarcinoma. Cancer Control. 2017;24(3):1073274817729245. doi:10.1177/1073274817729245

12. Fei FR, Hu RY, Gong WW, Pan J, Wang M. Analysis of Mortality and Survival Rate of

Liver Cancer in Zhejiang Province in China: A General Population-Based Study. Can J Gastroenterol Hepatol. 2019;2019:1074286. Published 2019 Jul 10. doi:10.1155/2019/1074286

13. Kim SY, Song HK, Lee SK, et al. Sex-Biased Molecular Signature for Overall Survival of Liver Cancer Patients. Biomol Ther (Seoul). 2020;28(6):491-502. doi:10.4062/biomolther.2020.157

14. Gabriele L., Buoncervello M., Ascione B., Bellenghi M., Matarrese P., Carè A. The gender perspective in cancer research and therapy: novel insights and on-going hypotheses. Ann. Ist. Super. Sanita. 2016;52:213–222. doi: 10.4415/ANN_16_02_13.

15. Liu K, Huang G, Chang P, et al. Construction and validation of a nomogram for predicting cancer-specific survival in hepatocellular carcinoma patients. Sci Rep. 2020;10(1):21376. Published 2020 Dec 7. doi:10.1038/s41598-020-78545-2

16. Zhang F, Lu S, Tian M, et al. Albumin-to-Alkaline Phosphatase Ratio is an Independent Prognostic Indicator in Combined Hepatocellular and Cholangiocarcinoma. J Cancer. 2020;11(17):5177-5186. Published 2020 Jun 29. doi:10.7150/jca.45633

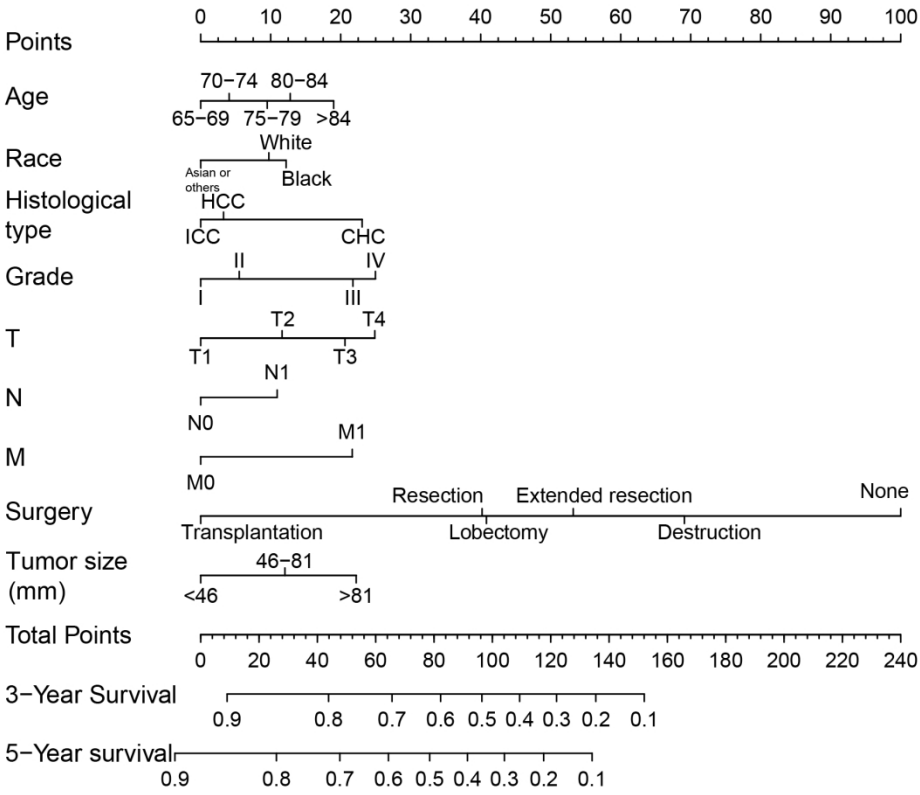
17. Ramai D, Ofosu A, Lai JK, Reddy M, Adler DG. Combined Hepatocellular Cholangiocarcinoma: A Population-Based Retrospective Study. Am J Gastroenterol. 2019;114(9):1496-1501. doi:10.14309/ajg.0000000000000326

18. Stavraka C, Rush H, Ross P. Combined hepatocellular cholan-giocarcinoma (c HCC-CC): an update of genetics, molecular biol-ogy, and therapeutic interventions. J Hepatocell Carcinoma,2019,6:11-21.

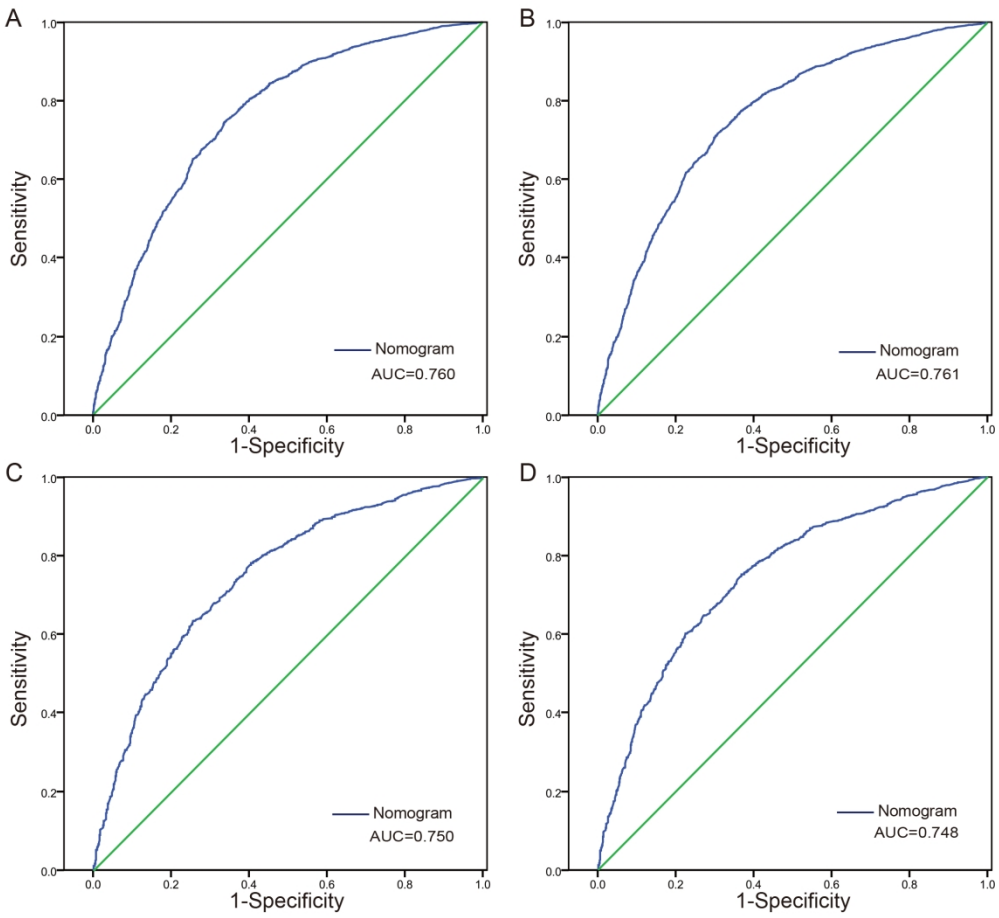
19. Leoni S, Sansone V, Lorenzo S, et al. Treatment of Combined Hepatocellular and Cholangiocarcinoma. Cancers (Basel). 2020;12(4):794. Published 2020 Mar 26. doi:10.3390/cancers12040794

20. EW, S. et al. Prediction of indolent prostate cancer: validation and updating of a prognostic nomogram. J. Urol. 177, 107–112; discussion 112, doi:https://doi.org/10.1016/j.juro.2006.08.068 (2007).

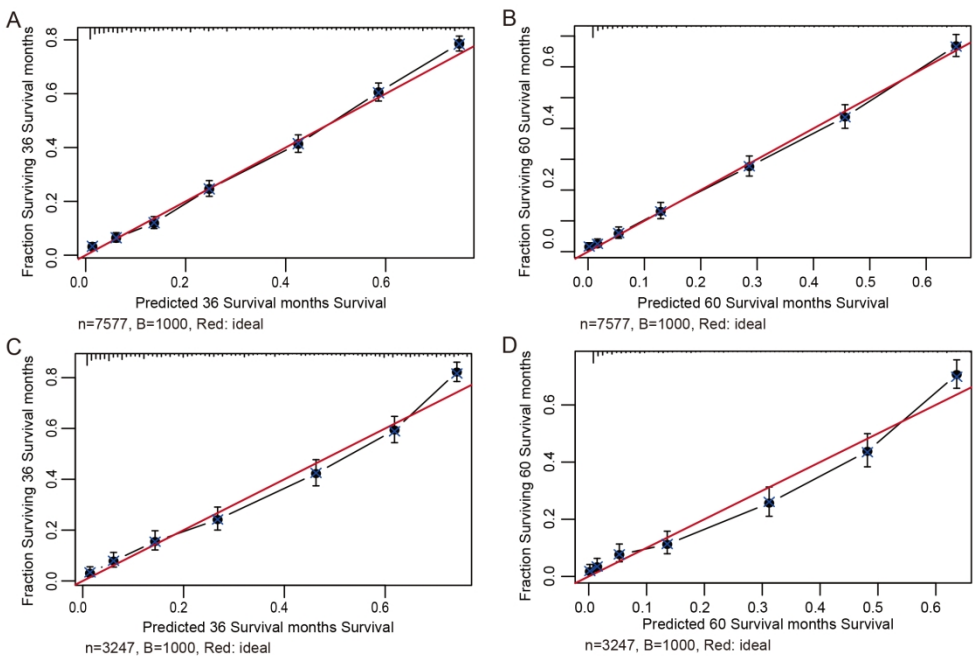
21. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol. 2008;26(8):1364-1370. doi:10.1200/JCO.2007.12.9791



170x136mm (300 x 300 DPI)



198x182mm (300 x 300 DPI)



188x125mm (300 x 300 DPI)

Table 1 Baseline data of extracted cases

Variates	Total cohort 10824(100%)	Training cohort 7577(100%)	Validation cohort 3247(100%)
Age			
65-69	3600(33.3%)	2540(33.5%)	1060(32.6%)
70-74	2829(26.1%)	1994(26.3%)	835(25.7%)
75-79	2228(20.6%)	1544(20.4%)	684(21.1%)
80-84	1451(13.4%)	1001(13.2%)	450(13.9%)
>84	716(6.61%)	498(6.57%)	218(6.71%)
Sex			
Male	7309(67.5%)	5122(67.6%)	2187(67.4%)
Female	3515(32.5%)	2455(32.4%)	1060(32.6%)
Race			
White	7722(71.3%)	5390(71.1%)	2332(71.8%)
Black	956(8.83%)	672(8.87%)	284(8.75%)
Asian or others	2146(19.8%)	1515(20.0%)	631(19.4%)
Primary site			
Liver	9508(87.8%)	6674(88.1%)	2834(87.3%)
IBD	1316(12.2%)	903(11.9%)	413(12.7%)
Histological type			
HCC	9171(84.7%)	6419(84.7%)	2752(84.8%)
ICC	1570(14.5%)	1095(14.5%)	475(14.6%)
CHC	83(0.77%)	63(0.83%)	20(0.62%)
Grade			
I	3108(28.7%)	2163(28.5%)	945(29.1%)
II	5040(46.6%)	3510(46.3%)	1530(47.1%)
III	2504(23.1%)	1785(23.6%)	719(22.1%)
IV	172(1.59%)	119(1.57%)	53(1.63%)
T			
T1	5028(46.5%)	3523(46.5%)	1505(46.4%)
T2	2547(23.5%)	1786(23.6%)	761(23.4%)
T3	2765(25.5%)	1932(25.5%)	833(25.7%)
T4	484(4.47%)	336(4.43%)	148(4.56%)
N			
N0	9910(91.6%)	6931(91.5%)	2979(91.7%)
N1	914(8.44%)	646(8.53%)	268(8.25%)
M			
M0	9605(88.7%)	6716(88.6%)	2889(89.0%)
M1	1219(11.3%)	861(11.4%)	358(11.0%)
Surgery			
Resection	1901(17.5%)	1315(17.4%)	586(18.0%)
Lobectomy	1116(10.3%)	807(10.7%)	309(9.52%)

Transplantation	328(3.03%)	238(3.14%)	90(2.77%)
Destruction	1087(10.0%)	769(10.1%)	318(9.79%)
Extended resection	277(2.56%)	195(2.56%)	82(2.53%)
None	6115(56.5%)	4253(56.1%)	1862(57.3%)
Tumor size(mm):			
<46	4168(38.5%)	2925(38.6%)	1243(38.3%)
46-81	3532(32.6%)	2491(32.9%)	1041(32.1%)
>81	3124(28.9%)	2161(28.5%)	963(29.7%)

Table 2 Univariate COX analysis

Variates	P-value	Hazard ratio	95%CI lower	95%CI upper
Age	**<0.001**			
65-69	Reference			
70-74	<0.001	1.168	1.085	1.257
75-79	<0.001	1.420	1.314	1.534
80-84	<0.001	1.639	1.502	1.788
>84	<0.001	2.072	1.855	2.314
Sex	**0.003**			
Male	Reference			
Female	0.003	0.914	0.862	0.969
Race	**<0.001**			
White	Reference			
Black	0.224	1.062	0.964	1.170
Asian or others	<0.001	0.737	0.685	0.792
Primary site	**0.232 (Excluded)**			
Liver	Reference			
IBD	0.232	1.055	0.967	1.151
Histological type	**0.032**			
HCC	Reference			
ICC	0.383	0.881	0.663	1.171
CHC	0.861	0.974	0.727	1.305
Grade	**<0.001**			
I	Reference			
II	0.043	0.934	0.875	0.998
III	<0.001	1.464	1.360	1.577
IV	0.001	1.437	1.162	1.776
T	**<0.001**			
T1	Reference			
T2	<0.001	1.213	1.129	1.304
T3	<0.001	2.446	2.290	2.614
T4	<0.001	2.493	2.200	2.825
N	**<0.001**			
N0	Reference			
N1	<0.001	2.265	2.072	2.476
M	**<0.001**			
M0	Reference			
M1	<0.001	3.025	2.798	3.271
Surgery	**<0.001**			
Resection	Reference			
Lobectomy	<0.001	0.234	0.213	0.256

Transplantation	<0.001	0.268	0.241	0.299
Destruction	<0.001	0.079	0.060	0.104
Extended resection	<0.001	0.366	0.332	0.403
None	<0.001	0.372	0.308	0.449
Tumor size(mm):	**<0.001**			
<46	Reference			
46-81	<0.001	1.744	1.630	1.867
>81	<0.001	2.577	2.405	2.761

CI: confidence interval.

Table 3 Multivariates COX analysis

Variates	P-value	Hazard ratio	95%CI lower	95%CI upper
Age				
65-69	Reference			
70-74	0.029	1.086	1.009	1.170
75-79	<0.001	1.219	1.127	1.318
80-84	<0.001	1.307	1.196	1.428
>84	<0.001	1.484	1.326	1.660
Sex	**(Excluded)**			
Male				
Female				
Race				
White	Reference			
Black	0.325	1.051	0.952	1.159
Asian or others	<0.001	0.813	0.756	0.875
Histological type				
HCC	Reference			
ICC	0.159	0.940	0.863	1.024
CHC	0.005	1.508	1.132	2.010
Grade				
I	Reference			
II	0.001	1.121	1.047	1.199
III	<0.001	1.567	1.449	1.695
IV	<0.001	1.683	1.358	2.086
T				
T1	Reference			
T2	<0.001	1.282	1.190	1.381
T3	<0.001	1.542	1.435	1.657
T4	<0.001	1.689	1.484	1.923
N				
N0	Reference			
N1	<0.001	1.253	1.136	1.382
M				
M0	Reference			
M1	<0.001	1.556	1.429	1.694
Surgery				
Resection	Reference			
Lobectomy	0.833	1.014	0.889	1.157
Transplantation	<0.001	0.417	0.313	0.557
Destruction	<0.001	1.851	1.632	2.100
Extended resection	0.007	1.325	1.080	1.626

None	<0.001	3.552	3.229	3.906
Tumor size(mm):				
<46	Reference			
46-81	<0.001	1.291	1.199	1.391
>81	<0.001	1.597	1.474	1.730

CI: confidence interval.

Figure 1 Constructed nomogram

Figure 2 3- and 5-year survival ROC curves for the training and validation cohorts. A: 3-year survival ROC curve for the training cohort. B: 5-year survival ROC curve for the training cohort. C: 3-year survival ROC curve for the validation cohort. D: 5-year survival ROC curve for the validation cohort.

Figure 3 3- and 5-year survival calibration curves for the training and validation cohorts. A: 3-year survival calibration curve for the training cohort. B: 5-year survival calibration curve for the training cohort. C: 3-year survival calibration curve for the validation cohort. D: 5-year survival calibration curve for the validation cohort.

STROBE 2007 (v4) checklist of items to be included in reports of observational studies in epidemiology*

Checklist for cohort, case-control, and cross-sectional studies (combined)

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any pre-specified hypotheses	3
Methods			
Study design	4	Present key elements of study design early in the paper	4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	4
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*4*
Bias	9	Describe any efforts to address potential sources of bias	4
Study size	10	Explain how the study size was arrived at	4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	4
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	5
		(*b*) Describe any methods used to examine subgroups and interactions	5
		(*c*) Explain how missing data were addressed	5
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed	5

		Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	
		(*e*) Describe any sensitivity analyses	
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	5
		(b) Give reasons for non-participation at each stage	
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	6
		(b) Indicate number of participants with missing data for each variable of interest	
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	6
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	6
		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	
Discussion			
Key results	18	Summarise key results with reference to study objectives	7
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	7
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	7
Generalisability	21	Discuss the generalisability (external validity) of the study results	7
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	8

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.
Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Prognostic risk factor analysis and construction of a nomogram for elderly primary liver cancer patients based on SEER database

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-051946.R1
Article Type:	Original research
Date Submitted by the Author:	13-May-2022
Complete List of Authors:	Li, Fangyuan; The first People's Hospital of Yuhang District, Department of Medical Oncology Zheng, Ting; The first People's Hospital of Yuhang District, Department of Medical Oncology Gu, Xuewei; Zhuji People's Hospital of Zhejiang Province, Department of Gastroenterology
Primary Subject Heading:	Oncology
Secondary Subject Heading:	Oncology
Keywords:	Cancer genetics < GENETICS, Adult oncology < ONCOLOGY, Gene therapy < ONCOLOGY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Prognostic risk factor analysis and construction of a nomogram for elderly primary liver cancer patients based on SEER database

Fangyuan Li[1], Ting Zheng[1], Xuewei Gu[2,*]

[1] Department of Medical Oncology, The first People's Hospital of Yuhang District, Hangzhou, 369 Yingbin Road,Yuhang District, Hangzhou, 311100, Zhejiang, PR China.

[2] Department of Gastroenterology, Zhuji People's Hospital of Zhejiang Province (Zhuji Affiliated Hospital of Shaoxing University), 9 Jianmin Road, Taozhu street, Zhuji, 311800, Zhejiang, China.

*Corresponding Author（Xuewei Gu）: Department of Gastroenterology, Zhuji People's Hospital of Zhejiang Province (Zhuji Affiliated Hospital of Shaoxing University), 9 Jianmin Road, Taozhu street, Zhuji, 311800, Zhejiang, China.

E-mail: gxwgxw1234567@163.com

Running title: Nomogram model for primary liver cancer in elderly patients

Key words: nomogram, primary liver cancer, elderly, database

Abstract

Objective: To evaluate the risk factors and construct a nomogram model for the prognosis of primary liver cancer in the elderly based on the data from the US SEER database. **Methods:** The latest data of primary liver cancer patients were extracted from the SEER database with SEER*STAT software, and the required variables were included. The data was screened and then divided into training cohort and validation cohort. Variates were first screened by univariate and multivariate COX analysis, and variates with then statistical significance were included the construction of a nomogram model. The C-Index, ROC and calibration curves were used to evaluate the model. **Results:** A total of 10824 eligible cases from 2004 to 2017 were extracted, among which, 7757 cases were included in the training cohort and 3247 in the validation cohort. The C-Index of the model was 0.747 (training cohort) and 0.773 (validation cohort). The 3-year AUCs of the training and validation cohort were 0.760 and 0.750, while the 5-year AUCs were 0.761 and 0.748. The calibration curves showed an ideal calibration of the constructed model. **Conclusions:** The nomogram model constructed followed by COX regression analysis showed ideal calibration and discrimination property, and can provide a reference for clinical application of elderly primary liver cancer in the future.

Strength and limitations of this study:

Strength:

1.Collected a large and sufficient number of cases from SEER database of elderly liver cancer.

2.Constructed a novel and ideal prognostic model for elderly liver cancer

Limitations:

1.All cases were collected form one database and selection bias might exist.

2.Some classification carried out by SEER database was not very specific.

3.Information such as ancillary tests was absent from the SEER database.

Keywords: nomogram, primary liver cancer, elderly, database

Introduction

Primary liver cancer is currently the sixth most common cancer worldwide and, according to epidemiological surveys, is the fourth leading cause of cancer-related deaths globally, representing a major threat to the health of the entire human population [1,2]. Furthermore, many studies have pointed out that although middle-aged (30-59 years old) or young (< 30 years old) patients with primary liver cancer are not uncommon worldwide, the average age of diagnosis of the disease is 60 years old and, in contrast to the yearly decrease in the age-standardized incidence rate (ASR) of young patients, the incidence of elderly patients has continuously increased in more than half of

the countries and regions during the last 30 years [3-5]. The global population expansion, increasing aging, as well as obesity, diabetes, overmedication and lagging effects of HBV infection in the elderly may be responsible for the high or even increasing ASR of primary liver cancer in elderly patients, creating a heavy burden on the health of all countries [6-8]. In terms of treatment, surgery remains the first choice for primary liver cancer. Therefore, based on the epidemiological characteristics and treatment modalities of the disease, it is necessary to conduct an accurate prognostic assessment of elderly patients with primary liver cancer to guide clinical work. However, the different pathological types and heterogeneity of the disease make prognostic assessment still difficult.

Recently, the nomogram model has shown superior predictive performance over the traditional TNM staging because of its convenient modeling method and the ability to incorporate multiple variables, thus gaining widespread popularity [9,10]. This study intended to use a nomogram model to analyze the risk factors affecting elderly primary liver cancer in the SEER database and to predict the prognosis of the disease. The assessment performance of the model was analyzed by the test of discrimination and calibration to establish an optimal assessment system for clinical work such as the treatment of elderly primary liver cancer.

Methods and data

1. Ethical statement

Informed consent was not required from patients to obtain data from the US SEER database since cancer is publicly reportable in every state in the United States.

2. Patient and Public Involvement

No patient involved.

3. Case selection

Case data of primary liver cancer with complete follow-up records were selected from the

2004-2017 SEER database (SEER research data, 18 Registries, Nov 2019 Sub (2000-2017)) using SEER*Stat 8.3.6.

Inclusion criteria:

1. Ethnic groups are Asian, Pacific Islanders, American Indians and Alaskans.

2. The main site of primary liver cancer is liver or intrahepatic bile duct (IBD).

3. The histological types of primary liver cancer are intrahepatic cholangiocarcinoma (ICC), hepatocellular carcinoma (HCC) and associated liver cancer (CHC).

Exclusion criteria:

1.For patients under 65 years old.

2.For incomplete follow-up records.

3. Non-tumor-related death.

The extracted variables included race, year of diagnosis, age, sex, primary site, histologic type, grade, TNM stage, tumor size, surgery on the primary site（included photodynamic therapy (PDT), percutaneous ethanol injection (PEI) and radiofrequency ablation (RFA), etc.）, survival time, cause of death and survival status. Among them, patients who were older than 65 years old were selected; Asian and Pacific Islanders, American Indians and Alaskan natives were included in Asian and others for the race variable; liver or intrahepatic bile duct (IBD) was selected as the primary site; intrahepatic cholangiocarcinoma (ICC), hepatocellular carcinoma (HCC), and combined hepatic carcinoma (CHC) were selected as the histologic type.

4. Statistical processing

The survival endpoint and survival time were defined as 3 years and 5 years, separately. Using X-Tile software, through the "enumeration method", that is, the statistical test is carried out by grouping different values as cut-off values, the result with the smallest p value can be considered as the best cut-off value. It is concluded that the variables of high, medium and low risk are divided into < 46mm, 46-81mm and > 81mm respectively.. After that, the total cases were randomly assigned to a training or a validation cohort at a ratio of 7:3 using SPSS 18.0 by random

number 20200222, followed by the collection of baseline information. Univariate and multivariate (Forward: LR) COX analyses were performed using the R software or SPSS to screen for statistically significant variables to construct a nomogram, based on which C-Index, ROC curves and the area under the curve (AUC) were figured out. Calibration curves of the model for 3 and 5 years were plotted with the R software after Bootstrap sampling for 1000 times. $P < 0.05$ was considered statistically significant.

Results

1. Clinical characteristics of the cases

A total of 10,824 elderly cases with primary liver cancer were extracted in accordance with screening conditions, with 7,757 included in the training cohort and 3,247 in the validation one. Among them, the majority of patients were male (67.5%), white (71.3%), with primary site in the liver (87.8%), HCC (84.7%), grade II (46.6%), T1 (46.5%), N0 (91.6%), M0 (88.7%) and unoperated (56.5%) (Table 1).

Table 1 Baseline data of extracted cases

Variates	Total cohort 10824(100%)	Training cohort 7577(100%)	Validation cohort 3247(100%)
Age			
65-69	3600(33.3%)	2540(33.5%)	1060(32.6%)
70-74	2829(26.1%)	1994(26.3%)	835(25.7%)
75-79	2228(20.6%)	1544(20.4%)	684(21.1%)
80-84	1451(13.4%)	1001(13.2%)	450(13.9%)
>84	716(6.61%)	498(6.57%)	218(6.71%)
Sex			
Male	7309(67.5%)	5122(67.6%)	2187(67.4%)

Female	3515(32.5%)	2455(32.4%)	1060(32.6%)
Race			
White	7722(71.3%)	5390(71.1%)	2332(71.8%)
Black	956(8.83%)	672(8.87%)	284(8.75%)
Asian or others	2146(19.8%)	1515(20.0%)	631(19.4%)
Primary site			
Liver	9508(87.8%)	6674(88.1%)	2834(87.3%)
IBD	1316(12.2%)	903(11.9%)	413(12.7%)
Histological type			
HCC	9171(84.7%)	6419(84.7%)	2752(84.8%)
ICC	1570(14.5%)	1095(14.5%)	475(14.6%)
CHC	83(0.77%)	63(0.83%)	20(0.62%)
Grade			
I	3108(28.7%)	2163(28.5%)	945(29.1%)
II	5040(46.6%)	3510(46.3%)	1530(47.1%)
III	2504(23.1%)	1785(23.6%)	719(22.1%)
IV	172(1.59%)	119(1.57%)	53(1.63%)
T			
T1	5028(46.5%)	3523(46.5%)	1505(46.4%)
T2	2547(23.5%)	1786(23.6%)	761(23.4%)
T3	2765(25.5%)	1932(25.5%)	833(25.7%)
T4	484(4.47%)	336(4.43%)	148(4.56%)
N			
N0	9910(91.6%)	6931(91.5%)	2979(91.7%)
N1	914(8.44%)	646(8.53%)	268(8.25%)
M			
M0	9605(88.7%)	6716(88.6%)	2889(89.0%)

M1	1219(11.3%)	861(11.4%)	358(11.0%)
Surgery			
Resection	1901(17.5%)	1315(17.4%)	586(18.0%)
Lobectomy	1116(10.3%)	807(10.7%)	309(9.52%)
Transplantation	328(3.03%)	238(3.14%)	90(2.77%)
Destruction	1087(10.0%)	769(10.1%)	318(9.79%)
Extended resection	277(2.56%)	195(2.56%)	82(2.53%)
None	6115(56.5%)	4253(56.1%)	1862(57.3%)
Tumor size(mm):			
<46	4168(38.5%)	2925(38.6%)	1243(38.3%)
46-81	3532(32.6%)	2491(32.9%)	1041(32.1%)
>81	3124(28.9%)	2161(28.5%)	963(29.7%)

2. Screening for prognostic risk factors

Univariate COX regression analysis of the training cohort proved the variates of age, sex, race, histological type, grade, TNM stage, surgery and tumor size to be statistically significant ($P < 0.05$), which were included in the follow-up multivariate COX analysis. However, the primary site was excluded according to the analysis ($P = 0.232$) (Table 2). Subsequently, the sex variable was further excluded from the experiment by Forward: LR multivariate COX (Table 3). In conclusion, age, race, histological type, grade, TNM stage, surgery, and tumor size were independent risk factors that affect the prognosis of elderly patients from the extracted data with primary liver cancer, which could be used to construct a subsequent nomogram prediction model.

Table 2 Univariate COX analysis

Variates	P-value	Hazard ratio	95%CI lower	95%CI upper
Age	**<0.001**			

65-69	Reference			
70-74	<0.001	1.168	1.085	1.257
75-79	<0.001	1.420	1.314	1.534
80-84	<0.001	1.639	1.502	1.788
>84	<0.001	2.072	1.855	2.314
Sex	**0.003**			
Male	Reference			
Female	0.003	0.914	0.862	0.969
Race	**<0.001**			
White	Reference			
Black	0.224	1.062	0.964	1.170
Asian or others	<0.001	0.737	0.685	0.792
Primary site	**0.232 (Excluded)**			
Liver	Reference			
IBD	0.232	1.055	0.967	1.151
Histological type	**0.032**			
HCC	Reference			
ICC	0.383	0.881	0.663	1.171
CHC	0.861	0.974	0.727	1.305
Grade	**<0.001**			
I	Reference			
II	0.043	0.934	0.875	0.998
III	<0.001	1.464	1.360	1.577
IV	0.001	1.437	1.162	1.776
T	**<0.001**			
T1	Reference			
T2	<0.001	1.213	1.129	1.304

T3	<0.001	2.446	2.290	2.614
T4	<0.001	2.493	2.200	2.825
N	**<0.001**			
N0	Reference			
N1	<0.001	2.265	2.072	2.476
M	**<0.001**			
M0	Reference			
M1	<0.001	3.025	2.798	3.271
Surgery	**<0.001**			
Resection	Reference			
Lobectomy	<0.001	0.234	0.213	0.256
Transplantation	<0.001	0.268	0.241	0.299
Destruction	<0.001	0.079	0.060	0.104
Extended resection	<0.001	0.366	0.332	0.403
None	<0.001	0.372	0.308	0.449
Tumor size(mm):	**<0.001**			
<46	Reference			
46-81	<0.001	1.744	1.630	1.867
>81	<0.001	2.577	2.405	2.761

CI: confidence interval.

Table 3 Multivariates COX analysis

Variates	P-value	Hazard ratio	95%CI lower	95%CI upper
Age				
65-69	Reference			
70-74	0.029	1.086	1.009	1.170

75-79	<0.001	1.219	1.127	1.318
80-84	<0.001	1.307	1.196	1.428
>84	<0.001	1.484	1.326	1.660
Sex	**(Excluded)**			
Male				
Female				
Race				
White	Reference			
Black	0.325	1.051	0.952	1.159
Asian or others	<0.001	0.813	0.756	0.875
Histological type				
HCC	Reference			
ICC	0.159	0.940	0.863	1.024
CHC	0.005	1.508	1.132	2.010
Grade				
I	Reference			
II	0.001	1.121	1.047	1.199
III	<0.001	1.567	1.449	1.695
IV	<0.001	1.683	1.358	2.086
T				
T1	Reference			
T2	<0.001	1.282	1.190	1.381
T3	<0.001	1.542	1.435	1.657
T4	<0.001	1.689	1.484	1.923
N				
N0	Reference			
N1	<0.001	1.253	1.136	1.382

M				
M0	Reference			
M1	<0.001	1.556	1.429	1.694
Surgery				
Resection	Reference			
Lobectomy	0.833	1.014	0.889	1.157
Transplantation	<0.001	0.417	0.313	0.557
Destruction	<0.001	1.851	1.632	2.100
Extended resection	0.007	1.325	1.080	1.626
None	<0.001	3.552	3.229	3.906
Tumor size(mm):				
<46	Reference			
46-81	<0.001	1.291	1.199	1.391
>81	<0.001	1.597	1.474	1.730

CI: confidence interval.

3. Nomogram model construction and verification

The independent risk factors affecting prognosis derived from the above analysis were incorporated to construct a 3- and 5-year nomogram prediction model for elderly primary liver cancer, and the total score was calculated by aggregating the scores of each variable to predict the survival rate of patients at 3 and 5 years (Figure 1). It can be seen that the surgery on the primary site was the most important factor affecting the score in this model, followed by tumor size, TNM stage and age. The AUC was calculated after plotting the ROC curves of the training and validation cohorts. Specifically, the AUC is 0.760 (3 years) and 0.761 (5 years) in the training cohort and 0.750 (3 years) and 0.748 (5 years) in the validation cohort (Figure 2). Furthermore, the model showed an ideal calibration for 3 and 5 years in both groups after creating the calibration curves for the training and validation cohorts (Figure 3). The comparison of predictive value between 3-years nomogram model and TNM model was with an AUC of 0.758 and 0.698 (P

<0.05) separately, and between 5-years nomogram model and TNM model was with an AUC of 0.750 and 0.609 (P <0.01), respectively (Figure 4.).

Discussion

Analysis of cases revealed that male patients accounted for more than 60% of elderly patients with primary liver cancer. Some statistics have presented that the mean annual change rate of men suffering from the disease is higher than that of women (3.7% vs. 2.7%) in the United States [11]. In China, a population-based study of hepatic carcinoma in Zhejiang Province demonstrated that the ASR for hepatic carcinoma was 33.24 in men compared to 1.21 in women [12]. Not only differences in lifestyle---including alcohol consumption and smoking---have led to higher cancer rates in men, but different physiological conditions such as hormone secretion and even genetic differences may be responsible for these epidemiological differences [13]. Therefore, it has been put forward that gender is a critical biological variable that should be considered in all studies aimed at improving carcinoma [14]. Analysis of baseline data also suggested that the population of elderly primary liver cancer was predominantly white and mostly with the primary site in the liver, HCC histological type, grade II (moderately differentiated), T1 and without lymph node metastasis or distant metastasis. Moreover, in this population, more than half of the cases were not treated surgically. The reason for this phenomenon may be that most of the patients were over 60 years old at the time of diagnosis, missing the best timing for radical surgery, together with the decline in physical function as well as intolerance to surgery led to a palliative treatment for most patients.

Based on further univariate and multivariate COX analyses, several independent risk factors that affect the prognosis of the disease were sifted out, including age, race, histological type, grade, TNM stage, surgery and tumor size. Sex, though not negligible as previously mentioned, was found not to be the main factor affecting prognosis in this population after comprehensive analysis, which is consistent with several current retrospective studies on hepatic carcinoma [15-17]. Some clinical information affecting the operation, such as metastatic cancer, can be

reflected in the TNM staging. In terms of histological type, it is evident that CHC has a worse prognosis than the common HCC, which shows a lower incidence but a higher malignancy [18-19]. Analysis of the age factor revealed that the higher the age group of the patient, the worse the prognosis, suggesting a linear negative correlation trend. The nomogram model also indicated that the surgery option was the most crucial factor influencing the prognosis of the disease. The patients who underwent liver transplantation, though small in number, showed a relatively good prognosis, followed by resection or lobectomy and then by local destruction. In contrast, patients without surgery showed a poor prognosis. This factor alone reduced the 3- and 5-year predicted survival rates to less than 50%, suggesting that the invention of new methods or enhanced surgery is still an urgent issue for improving the prognosis of elderly primary liver cancer. The influence of other factors on prognosis is basically in line with the current consensus that the worse the grade, the higher the T-stage, the occurrence of lymph node metastasis, the occurrence of distant metastasis and the larger the tumor, the worse the prognosis of the patient.

After that, the performance of the established model was evaluated by C-Index, ROC curves and calibration curves. A nomogram model is considered to have good discrimination if its C-Index and AUC exceed 0.7 [20,21]. As the model constructed in this study had these two indicators above 0.7 in both the training and validation cohorts and the calibration plots scatter in accordance with the reference line, it could be considered that the model has good discrimination and calibration and hence the capacity to predict the prognosis of the disease.

However, this study also has shortcomings. First, the cases were all from the US SEER database, which is not representative of regions other than the United States and is subject to selection bias. In addition, the case data included in this database lacked some important ancillary tests related to the diagnosis and treatment of liver cancer, such as CEA, AST and vascular invasion.More importantly, the radiotherapy and chemotherapy information contained in this database can only be obtained by signing some agreements, which can not be obtained for the time being, so we are unable to study the relationship between radiotherapy, chemotherapy, targeted therapy and the prognosis of liver cancer [22].

In conclusion, a nomogram model with favorable prediction was developed by using the case data from the SEER database after performing univariate and multivariate COX screening,

which could provide a reference for the future diagnosis and treatment of elderly patients with primary liver cancer.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Conflicts of interest

The authors declare there are no competing interests.

Author contributions

Fangyuan Li wrote and revised the manuscript; Ting Zheng conducted most of the analysis of data; Xuewei Gu reviewed the manuscript; All authors read and approved the final manuscript.

Availability of data and material

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Informed consent was not required from patients to obtain data from the US SEER database since cancer is publicly reportable in every state in the United States.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;70(1):7-30. doi:10.3322/caac.21590

2. Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. 68, 394–424. https ://doi. org/10.3322/caac.21492 (2018).

3. Liu Z, Suo C, Mao X, et al. Global incidence trends in primary liver cancer by age at diagnosis, sex, region, and etiology, 1990-2017. Cancer. 2020;126(10):2267-2278. doi:10.1002 /cncr.32789

4. Kanwal F, Kramer JR, Mapakshi S, et al. Risk of hepatocellular cancer in patients with non-alcoholic fatty liver disease. Gastroenterology. 2018;155:1828-1837.e2.

5. Zhang S, Zheng R, Zeng H, Chen W. The incidence differences among sex, geographical areas and mean age of diagnosis for liver cancer in China, 1989-2008 [in Chinese]. Zhonghua Yu Fang Yi Xue Za Zhi. 2014;48:355-360.

6. Baumeister SE, Schlesinger S, Aleksandrova K, et al. Association between physical activity and risk of hepatobiliary cancers: a multinational cohort study. J Hepatol. 2019;70:885-892.

7. Klil-Drori AJ, Azoulay L, Pollak MN. Cancer, obesity, diabetes, and antidiabetic drugs: is the fog clearing? Nat Rev Clin Oncol. 2017;14:85-99.

8. Smittenaar CR, Petersen KA, Stewart K, Moitt N. Cancer incidence and mortality projections in the UK until 2035. Br J Cancer. 2016;115(9):1147-1155. doi:10.1038/bjc.2016.304

9. Chen, S., et al., Development and Validation of a Nomogram for Predicting Survival in Male Patients With Breast Cancer. Front Oncol, 2019. 9: p. 361

10. Zhang Y, Hong YK, Zhuang DW, He XJ, Lin ME. Bladder cancer survival nomogram: Development and validation of a prediction tool, using the SEER and TCGA databases. Medicine (Baltimore). 2019 Nov;98(44):e17725.

11. Massarweh NN, El-Serag HB. Epidemiology of Hepatocellular Carcinoma and Intrahepatic Cholangiocarcinoma. Cancer Control. 2017;24(3):1073274817729245. doi:10.1177/1073274817729245

12. Fei FR, Hu RY, Gong WW, Pan J, Wang M. Analysis of Mortality and Survival Rate of Liver Cancer in Zhejiang Province in China: A General Population-Based Study. Can J Gastroenterol Hepatol. 2019;2019:1074286. Published 2019 Jul 10. doi:10.1155/2019/1074286

13. Kim SY, Song HK, Lee SK, et al. Sex-Biased Molecular Signature for Overall Survival of Liver Cancer Patients. Biomol Ther (Seoul). 2020;28(6):491-502. doi:10.4062/biomolther.2020.157

14. Gabriele L., Buoncervello M., Ascione B., Bellenghi M., Matarrese P., Carè A. The gender perspective in cancer research and therapy: novel insights and on-going hypotheses. Ann. Ist. Super. Sanita. 2016;52:213–222. doi: 10.4415/ANN_16_02_13.

15. Liu K, Huang G, Chang P, et al. Construction and validation of a nomogram for predicting cancer-specific survival in hepatocellular carcinoma patients. Sci Rep. 2020;10(1):21376. Published 2020 Dec 7. doi:10.1038/s41598-020-78545-2

16. Zhang F, Lu S, Tian M, et al. Albumin-to-Alkaline Phosphatase Ratio is an Independent Prognostic Indicator in Combined Hepatocellular and Cholangiocarcinoma. J Cancer. 2020;11(17):5177-5186. Published 2020 Jun 29. doi:10.7150/jca.45633

17. Ramai D, Ofosu A, Lai JK, Reddy M, Adler DG. Combined Hepatocellular Cholangiocarcinoma: A Population-Based Retrospective Study. Am J Gastroenterol. 2019;114(9):1496-1501. doi:10.14309/ajg.0000000000000326

18. Stavraka C, Rush H, Ross P. Combined hepatocellular cholan-giocarcinoma (c HCC-CC): an update of genetics, molecular biol-ogy, and therapeutic interventions. J Hepatocell

Carcinoma,2019,6:11-21.

19. Leoni S, Sansone V, Lorenzo S, et al. Treatment of Combined Hepatocellular and Cholangiocarcinoma. Cancers (Basel). 2020;12(4):794. Published 2020 Mar 26. doi:10.3390/cancers12040794

20. EW, S. et al. Prediction of indolent prostate cancer: validation and updating of a prognostic nomogram. J. Urol. 177, 107–112; discussion 112, doi:https://doi.org/10.1016/j.juro.2006.08.068 (2007).

21. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol. 2008;26(8):1364-1370. doi:10.1200/JCO.2007.12.9791

22. Yang J, Li YJ, Liu QQ, Li L, Feng AZ, Wang TY, Zheng S, Xu AD, Lyu J. Brief introduction of medical database and data mining technology in big data era. Journal of Evidence-Based Medicine 2020, 13: 57-69.

Figure Legends

Figure 1 Constructed nomogram

Figure 2 3- and 5-year survival ROC curves for the training and validation cohorts. A: 3-year survival ROC curve for the training cohort. B: 5-year survival ROC curve for the training cohort. C: 3-year survival ROC curve for the validation cohort. D: 5-year survival ROC curve for the validation cohort.

Figure 3 3- and 5-year survival calibration curves for the training and validation cohorts. A: 3-year survival calibration curve for the training cohort. B: 5-year survival calibration curve for the training cohort. C: 3-year survival calibration curve for the validation cohort. D: 5-year survival calibration curve for the validation cohort.

Figure.4 The comparison of ROC between nomogram model and TNM model. (A: 3-years nomogram model，B: 5-years nomogram model)

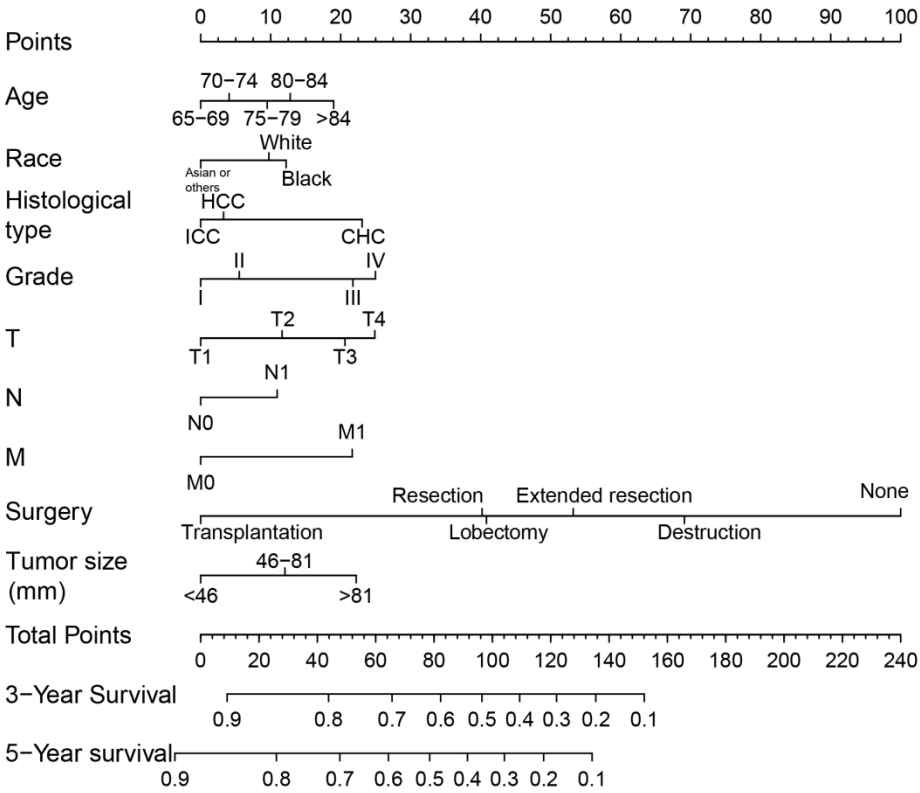


Figure 1 Constructed nomogram

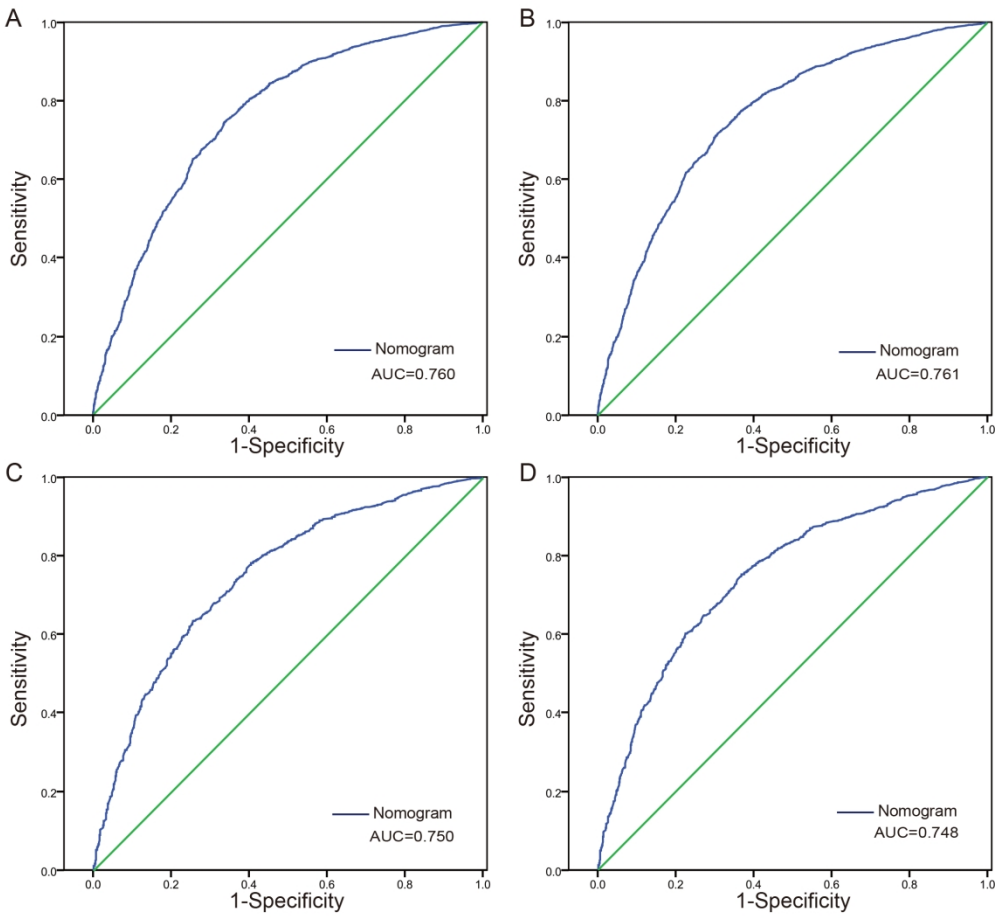


Figure 2 3- and 5-year survival ROC curves for the training and validation cohorts. A: 3-year survival ROC curve for the training cohort. B: 5-year survival ROC curve for the training cohort. C: 3-year survival ROC curve for the validation cohort. D: 5-year survival ROC curve for the validation cohort.

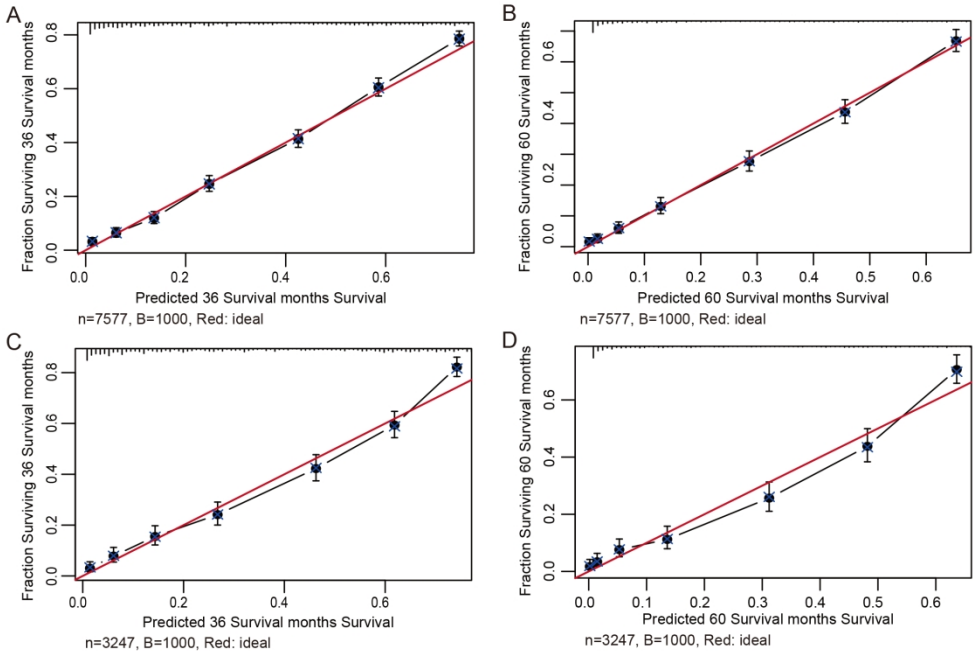


Figure 3 3- and 5-year survival calibration curves for the training and validation cohorts. A: 3-year survival calibration curve for the training cohort. B: 5-year survival calibration curve for the training cohort. C: 3-year survival calibration curve for the validation cohort. D: 5-year survival calibration curve for the validation cohort.

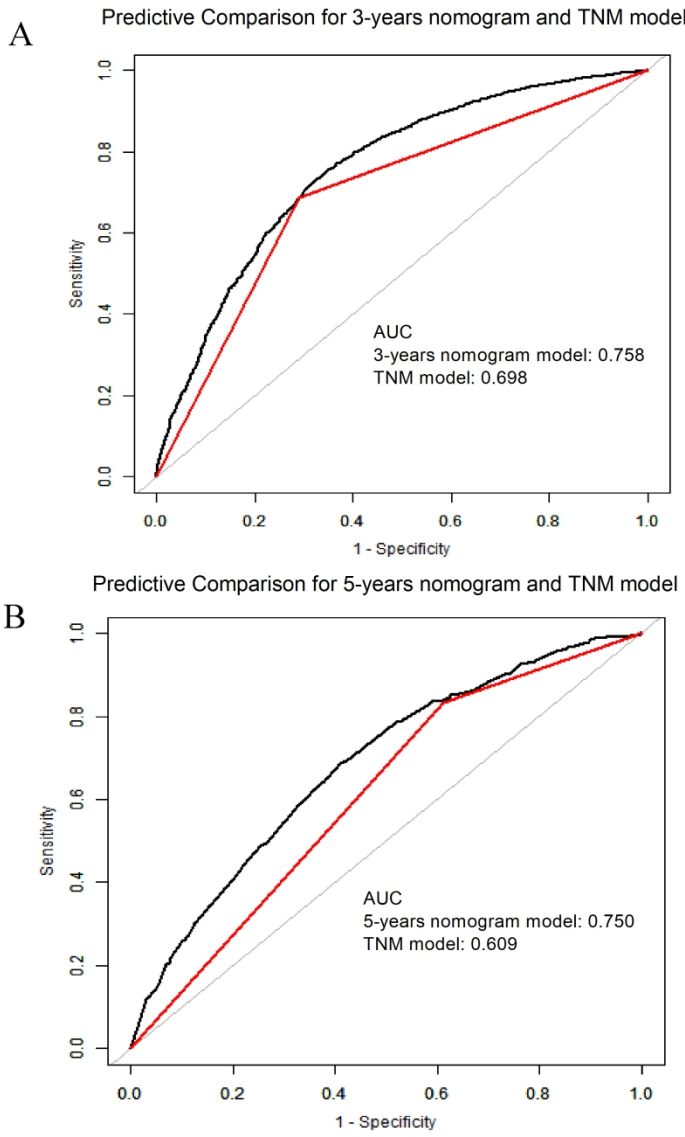


Figure.4 The comparison of ROC between nomogram model and TNM model. (A: 3-years nomogram model, B: 5-years nomogram model)

203x291mm (300 x 300 DPI)

STROBE 2007 (v4) checklist of items to be included in reports of observational studies in epidemiology*

Checklist for cohort, case-control, and cross-sectional studies (combined)

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any pre-specified hypotheses	3
Methods			
Study design	4	Present key elements of study design early in the paper	4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	4
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*4*
Bias	9	Describe any efforts to address potential sources of bias	4
Study size	10	Explain how the study size was arrived at	4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	4
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	5
		(*b*) Describe any methods used to examine subgroups and interactions	5
		(*c*) Explain how missing data were addressed	5
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed	5

		Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	
		(*e*) Describe any sensitivity analyses	
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	5
		(b) Give reasons for non-participation at each stage	
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	6
		(b) Indicate number of participants with missing data for each variable of interest	
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	6
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	6
		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	
Discussion			
Key results	18	Summarise key results with reference to study objectives	7
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	7
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	7
Generalisability	21	Discuss the generalisability (external validity) of the study results	7
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	8

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.
Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Prognostic risk factor analysis and nomogram construction for primary liver cancer in elderly patients based on SEER database

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-051946.R2
Article Type:	Original research
Date Submitted by the Author:	05-Sep-2022
Complete List of Authors:	Li, Fangyuan; The first People's Hospital of Yuhang District, Department of Medical Oncology Zheng, Ting; The first People's Hospital of Yuhang District, Department of Medical Oncology Gu, Xuewei; Zhuji People's Hospital of Zhejiang Province, Department of Gastroenterology
Primary Subject Heading:	Oncology
Secondary Subject Heading:	Oncology
Keywords:	Cancer genetics < GENETICS, Adult oncology < ONCOLOGY, Gene therapy < ONCOLOGY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Prognostic risk factor analysis and nomogram construction for primary liver cancer in elderly patients based on SEER database

Fangyuan Li[1], Ting Zheng[1], Xuewei Gu[2,*]

[1] Department of Medical Oncology, The first People's Hospital of Yuhang District, Hangzhou, 369 Yingbin Road,Yuhang District, Hangzhou, 311100, Zhejiang, PR China.

[2] Department of Gastroenterology, Zhuji People's Hospital of Zhejiang Province (Zhuji Affiliated Hospital of Shaoxing University), 9 Jianmin Road, Taozhu street, Zhuji, 311800, Zhejiang, China.

*Corresponding Author (Xuewei Gu): Department of Gastroenterology, Zhuji People's Hospital of Zhejiang Province (Zhuji Affiliated Hospital of Shaoxing University), 9 Jianmin Road, Taozhu street, Zhuji, 311800, Zhejiang, China.

E-mail: gxwgxw1234567@163.com

Running title: Nomogram model for primary liver cancer in elderly patients

Key words: nomogram, primary liver cancer, elderly, database

Abstract

Objective: To evaluate the risk factors and construct a nomogram model for the prognosis of primary liver cancer in the elderly based on the data from the US SEER database. **Methods:** The latest data of patients with primary liver cancer were extracted from the SEER database using SEER*STAT software, and the required variables were included. The data were screened and then divided into a training cohort and a validation cohort. A nomogram model was constructed by screening the variables through univariate and multivariate COX analysis. The C-Index, ROC, and calibration curves were used for model evaluation. **Results:** A total of 10824 eligible cases from 2004 to 2017 were extracted, among which, 7757 cases were included in the training cohort and 3247 in the validation cohort. The C-Index of the model was 0.747 (in the training cohort) and 0.773 (in the validation cohort). The 3-year AUCs of the training and the validation cohorts were 0.760 and 0.750, and the 5-year AUCs of the two cohorts were 0.761 and 0.748. The calibration curves showed an ideal calibration of the constructed model. **Conclusions:** The nomogram model constructed followed by COX regression analysis showed moderate calibration and discrimination property, and can provide reference to a certain extent for furture clinical application of primary liver cancer in the elderly.

Strengths and limitations of this study:

Strengths:

1. A large and sufficient number of elderly cases with liver cancer were collected from the SEER database.

2. A novel and ideal prognostic model was constructed for the elderly patients with liver cancer.

Limitations:

1. Selection bias might exist, because all the cases were retrived from the same database.

2. Some of the classifications carried out in the SEER database were not specific enough.

3. Information such as ancillary tests were absent from the SEER database.

Keywords: nomogram, primary liver cancer, elderly, database

Introduction

Primary liver cancer is currently the sixth most common cancer worldwide, and is the fourth leading cause of cancer-related deaths globally according to epidemiological surveys, posing a major threat to the health of the entire human population [1,2]. Furthermore, many studies have pointed out that although middle-aged (30-59 years old) or young (< 30 years old) patients with primary liver cancer are not uncommon worldwide, the average age of diagnosis of the disease is 60. Besides, in contrast to the yearly decrease of the age-standardized incidence rate (ASR) among young patients, the incidence in elderly patients has continuously increased in more than half of the

countries and regions during the last 30 years [3-5]. Global population expansion, increasing aging, as well as obesity, diabetes, overmedication and lagging effects of HBV infection in the elderly may be responsible for the high or even increased ASR in elderly patients with primary liver cancer, imposing a heavy burden on the health sectors of all countries [6-8]. Surgery remains the first choice for the treatment of primary liver cancer. Therefore, based on the epidemiological characteristics and treatment modalities of primary liver cancer, it is necessary to accurately assess the prognosis of the disease in elderly patients for the guide of clinical practice. However, different pathological types and heterogeneity of the disease still make its prognostic assessment difficult.

Recently, the nomogram model has gained widespread popularity due to its superior predictive performance over the traditional TNM staging in the aspects of its convenient modeling method and ability to incorporate multiple variables [9,10]. This study intended to construct a nomogram model to analyze the risk factors of primary liver cancer in elderly patients base on the SEER database and to predict the prognosis of the disease. The evaluation effect of the model was analyzed by the test of discrimination and calibration, through which an optimal assessment system was established for the clinical practice such as the treatment of elderly patients with primary liver cancer.

Methods and data

1. Ethical statement

Informed consent was not required from patients to obtain data from the US SEER database since cancer is publicly reportable in every state in the United States.

2. Patient and public involvement

No patient involved.

3. Case selection

Case data of primary liver cancer with complete follow-up records were selected from the 2004-2017 SEER database (SEER research data, 18 Registries, Nov 2019 Sub (2000-2017)) using

SEER*Stat 8.3.6.

Inclusion criteria:

1. Ethnic groups are Asians, Pacific Islanders, American Indians and Alaskans.

2. The main site of primary liver cancer is liver or intrahepatic bile duct (IBD).

3. The histological types of primary liver cancer are intrahepatic cholangiocarcinoma (ICC), hepatocellular carcinoma (HCC) and associated liver cancer (CHC).

Exclusion criteria:

1.For patients under 65 years old.

2.For incomplete follow-up records.

3. Non-tumor-related death.

Race, year of diagnosis, age, sex, primary site, histologic type, grade, TNM stage, tumor size, surgery on the primary site (including photodynamic therapy (PDT), percutaneous ethanol injection (PEI) and radiofrequency ablation (RFA), etc.), survival time, cause of death and survival status were all extracted variables. Among them, patients over 65 years old were selected; Asians and Pacific Islanders, American Indians and Alaskan natives were included as the race variable of Asians and others; liver or intrahepatic bile duct (IBD) was selected as the primary site; intrahepatic cholangiocarcinoma (ICC), hepatocellular carcinoma (HCC), and combined hepatic carcinoma (CHC) were selected as the histologic type.

4. Statistical processing

The survival endpoint and survival time were defined as 3 years and 5 years, separately. The statistical test is carried out by grouping different values as cut-off values through the “enumeration method” using X-Tile software, and the result with the smallest p value can be considered as the best cut-off value. It was concluded that the variables of high, medium and low risks are divided into < 46mm, 46-81mm and > 81mm respectively. After that, all the cases were randomly assigned to a training or a validation cohort at a ratio of 7:3 using SPSS 18.0 by random number 20200222, followed by the collection of baseline information. Univariate and multivariate (Forward: LR) COX

analyses were performed using the R software or SPSS to screen statistically significant variables for nomogram construction, based on which, C-Index, ROC curves and the area under the curve (AUC) were figured out. Calibration curves of the model for 3 and 5 years were plotted with the R software after Bootstrap sampling for 1000 times. $P < 0.05$ was considered statistically significant.

Results

1. Clinical characteristics of the cases

A total of 10,824 elderly cases with primary liver cancer were extracted in accordance with the screening conditions, including 7,757 in the training cohort and 3,247 in the validation one. Among them, the majority of patients were male (67.5%), white (71.3%), with primary site in the liver (87.8%), HCC (84.7%), grade II (46.6%), T1 (46.5%), N0 (91.6%), M0 (88.7%) and unoperated (56.5%) (Table 1).

Table 1 Baseline data of the extracted cases

Variates	Total cohort 10824(100%)	Training cohort 7577(100%)	Validation cohort 3247(100%)
Age			
65-69	3600(33.3%)	2540(33.5%)	1060(32.6%)
70-74	2829(26.1%)	1994(26.3%)	835(25.7%)
75-79	2228(20.6%)	1544(20.4%)	684(21.1%)
80-84	1451(13.4%)	1001(13.2%)	450(13.9%)
>84	716(6.61%)	498(6.57%)	218(6.71%)
Sex			
Male	7309(67.5%)	5122(67.6%)	2187(67.4%)
Female	3515(32.5%)	2455(32.4%)	1060(32.6%)
Race			
White	7722(71.3%)	5390(71.1%)	2332(71.8%)
Black	956(8.83%)	672(8.87%)	284(8.75%)
Asian or others	2146(19.8%)	1515(20.0%)	631(19.4%)
Primary site			
Liver	9508(87.8%)	6674(88.1%)	2834(87.3%)

IBD	1316(12.2%)	903(11.9%)	413(12.7%)
Histological type			
HCC	9171(84.7%)	6419(84.7%)	2752(84.8%)
ICC	1570(14.5%)	1095(14.5%)	475(14.6%)
CHC	83(0.77%)	63(0.83%)	20(0.62%)
Grade			
I	3108(28.7%)	2163(28.5%)	945(29.1%)
II	5040(46.6%)	3510(46.3%)	1530(47.1%)
III	2504(23.1%)	1785(23.6%)	719(22.1%)
IV	172(1.59%)	119(1.57%)	53(1.63%)
T			
T1	5028(46.5%)	3523(46.5%)	1505(46.4%)
T2	2547(23.5%)	1786(23.6%)	761(23.4%)
T3	2765(25.5%)	1932(25.5%)	833(25.7%)
T4	484(4.47%)	336(4.43%)	148(4.56%)
N			
N0	9910(91.6%)	6931(91.5%)	2979(91.7%)
N1	914(8.44%)	646(8.53%)	268(8.25%)
M			
M0	9605(88.7%)	6716(88.6%)	2889(89.0%)
M1	1219(11.3%)	861(11.4%)	358(11.0%)
Surgery			
Resection	1901(17.5%)	1315(17.4%)	586(18.0%)
Lobectomy	1116(10.3%)	807(10.7%)	309(9.52%)
Transplantation	328(3.03%)	238(3.14%)	90(2.77%)
Destruction	1087(10.0%)	769(10.1%)	318(9.79%)
Extended resection	277(2.56%)	195(2.56%)	82(2.53%)
None	6115(56.5%)	4253(56.1%)	1862(57.3%)
Tumor size(mm):			
<46	4168(38.5%)	2925(38.6%)	1243(38.3%)
46-81	3532(32.6%)	2491(32.9%)	1041(32.1%)
>81	3124(28.9%)	2161(28.5%)	963(29.7%)

2. Screening for prognostic risk factors

Univariate COX regression analysis were performed on the training cohort, and the variates of age, sex, race, histological type, grade, TNM stage, surgery and tumor size were proved to be

statistically significant ($P < 0.05$) and included in the follow-up multivariate COX analysis. However, the primary site was excluded according to the analysis ($P = 0.232$) (Table 2). Subsequently, the variable of sex was further excluded from the experiment by Forward: LR multivariate COX (Table 3). In the end, age, race, histological type, grade, TNM stage, surgery, and tumor size were all independent risk factors affecting the prognosis of elderly patients with primary liver cancer, and could be used for constructing nomogram prediction model.

Table 2 Univariate COX analysis

Variates	P-value	Hazard ratio	95%CI lower	95%CI upper
Age	**<0.001**			
65-69	Reference			
70-74	<0.001	1.168	1.085	1.257
75-79	<0.001	1.420	1.314	1.534
80-84	<0.001	1.639	1.502	1.788
>84	<0.001	2.072	1.855	2.314
Sex	**0.003**			
Male	Reference			
Female	0.003	0.914	0.862	0.969
Race	**<0.001**			
White	Reference			
Black	0.224	1.062	0.964	1.170
Asian or others	<0.001	0.737	0.685	0.792
Primary site	**0.232 (Excluded)**			
Liver	Reference			
IBD	0.232	1.055	0.967	1.151
Histological type	**0.032**			
HCC	Reference			
ICC	0.383	0.881	0.663	1.171
CHC	0.861	0.974	0.727	1.305
Grade	**<0.001**			
I	Reference			
II	0.043	0.934	0.875	0.998
III	<0.001	1.464	1.360	1.577
IV	0.001	1.437	1.162	1.776
T	**<0.001**			

T1	Reference			
T2	<0.001	1.213	1.129	1.304
T3	<0.001	2.446	2.290	2.614
T4	<0.001	2.493	2.200	2.825
N	**<0.001**			
N0	Reference			
N1	<0.001	2.265	2.072	2.476
M	**<0.001**			
M0	Reference			
M1	<0.001	3.025	2.798	3.271
Surgery	**<0.001**			
Resection	Reference			
Lobectomy	<0.001	0.234	0.213	0.256
Transplantation	<0.001	0.268	0.241	0.299
Destruction	<0.001	0.079	0.060	0.104
Extended resection	<0.001	0.366	0.332	0.403
None	<0.001	0.372	0.308	0.449
Tumor size(mm):	**<0.001**			
<46	Reference			
46-81	<0.001	1.744	1.630	1.867
>81	<0.001	2.577	2.405	2.761

CI: confidence interval.

Table 3 Multivariates COX analysis

Variates	P-value	Hazard ratio	95%CI lower	95%CI upper
Age				
65-69	Reference			
70-74	0.029	1.086	1.009	1.170
75-79	<0.001	1.219	1.127	1.318
80-84	<0.001	1.307	1.196	1.428
>84	<0.001	1.484	1.326	1.660
Sex	**(Excluded)**			
Male				
Female				
Race				
White	Reference			

Black	0.325	1.051	0.952	1.159
Asian or others	<0.001	0.813	0.756	0.875
Histological type				
HCC	Reference			
ICC	0.159	0.940	0.863	1.024
CHC	0.005	1.508	1.132	2.010
Grade				
I	Reference			
II	0.001	1.121	1.047	1.199
III	<0.001	1.567	1.449	1.695
IV	<0.001	1.683	1.358	2.086
T				
T1	Reference			
T2	<0.001	1.282	1.190	1.381
T3	<0.001	1.542	1.435	1.657
T4	<0.001	1.689	1.484	1.923
N				
N0	Reference			
N1	<0.001	1.253	1.136	1.382
M				
M0	Reference			
M1	<0.001	1.556	1.429	1.694
Surgery				
Resection	Reference			
Lobectomy	0.833	1.014	0.889	1.157
Transplantation	<0.001	0.417	0.313	0.557
Destruction	<0.001	1.851	1.632	2.100
Extended resection	0.007	1.325	1.080	1.626
None	<0.001	3.552	3.229	3.906
Tumor size(mm):				
<46	Reference			
46-81	<0.001	1.291	1.199	1.391
>81	<0.001	1.597	1.474	1.730

CI: confidence interval.

3. Nomogram model construction and verification

The 3- and 5-year nomogram prediction model for primary liver cancer in the early were constructed based on the independent risk factors affecting the prognosis of the disease derived from the above analysis. The total score was calculated by aggregating the scores of each variable to

predict the 3- and 5- year survival rate of patients (Figure 1). It can be seen that the most important factor affecting the score in this model was surgery on the primary site, followed by tumor size, TNM stage and age. The AUC was calculated after plotting the ROC curves of the training and the validation cohorts. Specifically, the AUC is 0.760 (3 years) and 0.761 (5 years) in the training cohort, and 0.750 (3 years) and 0.748 (5 years) in the validation cohort (Figure 2). Furthermore, the model showed an ideal calibration for 3- and 5- year survival prediction in both groups after creating the calibration curves for the training and the validation cohorts (Figure 3). By comapring the predictive value of the nomogram model with the TNM model, it was revealed that their 3-yearAUC were 0.758 and 0.698 ($P < 0.05$) separately, and their 5-year AUC were 0.750 and 0.609 ($P < 0.01$), respectively (Figure 4.).

Discussion

Analysis of cases revealed that male patients accounted for more than 60% of all the elderly patients with primary liver cancer. Some statistics have presented that the mean annual change rate of men suffering from the disease is higher than that of women (3.7% vs. 2.7%) in the United States [11]. In China, a population-based study of hepatic carcinoma in Zhejiang Province demonstrated that the ASR for hepatic carcinoma was 33.24 in men compared to 1.21 in women [12]. Not only differences in lifestyle---including alcohol consumption and smoking---have led to higher cancer rates in men, but different physiological conditions such as hormone secretion and even genetic differences may be responsible for these epidemiological differences [13]. Therefore, it has been proposed that gender is a critical biological variable that should be considered in all studies aimed at improving carcinoma [14]. Analysis of baseline data also suggested that the population of elderly patients with primary liver cancer was predominantly white and mostly with the primary site in the liver, HCC histological type, grade II (moderately differentiated), T1 and without lymph node metastasis or distant metastasis. Moreover, in this population, more than half of the cases were not treated surgically. The possible reason for this phenomenon is that most of the patients were over 60 years old at the time of diagnosis, missing the best time to receive radical surgery. In addition,

in consideration of the decline in their physical function as well as intolerance to surgery, a palliative treatment was chosen for most of these patients.

Based on further univariate and multivariate COX analyses, several independent risk factors affecting the prognosis of the disease were obtained, including age, race, histological type, grade, TNM stage, surgery and tumor size. Sex, though not negligible as previously mentioned, was not a main factor affecting prognosis in this population after comprehensive analysis, which is consistent with several current retrospective studies on hepatic carcinoma [15-17]. Some clinical information affecting the operation, such as metastatic cancer, can be reflected in the TNM staging. In terms of histological types, the prognosis of CHC is obviously worse than that of the common HCC, with a lower incidence but a higher degree of malignancy [18-19]. Analysis of the age factor revealed that the higher the age group of the patient, the worse the prognosis, suggesting a linear negative correlation trend. The nomogram model also indicated that surgery was the most crucial factor influencing the prognosis of the disease. Although just a small number of patients received liver transplantation, they showed a relatively good prognosis, followed by patients with resection or lobectomy and local destruction. In contrast, patients without surgery showed a relatively poor prognosis. This factor alone reduced the 3- and 5-year predicted survival rates to less than 50%, suggesting that the invention of new methods or enhanced surgery is still urgent for improving the prognosis of elderly patients with primary liver cancer. The influence of other factors on the prognosis of the disease is basically in line with the current consensus that the worse the grade, the higher the T-stage, the occurrence of lymph node metastasis, the occurrence of distant metastasis and the larger the tumor, the worse the prognosis of the patients.

After that, the performance of the established model was evaluated by C-Index, ROC curves and calibration curves. A nomogram model is considered to have good discrimination if its C-Index and AUC exceed 0.7 [20,21]. As the two indicators of the model constructesd in this study were all above 0.7 in both the training and the validation cohorts and the calibration plots scattered in accordance with the reference line, it could be concluded that the model has good discrimination and calibration and hence the capacity to predict the prognosis of the disease.

However, this study also has shortcomings. First, the cases in this study were all from the US SEER database, which is not representative for regions outside the United States and is subject to

selection bias. In addition, the case data included in this database lacked some important ancillary tests related to the diagnosis and treatment of liver cancer, such as CEA, AST and vascular invasion. More importantly, the radiotherapy and chemotherapy information contained in this database can only be obtained by signing some agreements, which can not be obtained for the time being, so we are unable to study the relationship between radiotherapy, chemotherapy, targeted therapy and the prognosis of liver cancer [22].

There are also deficiencies in our statistical conclusions. Limited by time and skills, our model did not reach an ideal state, and its AUC is only 0.75, indicating that there is still room for improvement. This affects the prediction accuracy to a certain extent and reduces the prediction credibility. In the future, we will continue to refine our nomogram model to make it achieve a more accurate degree.

In conclusion, a nomogram model with moderate prediction was developed by using the case data in the SEER database after performing univariate and multivariate COX screening, which could provide reference for future diagnosis and treatment of elderly patients with primary liver cancer.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Conflicts of interest

The authors declare there are no competing interests.

Author contributions

Fangyuan Li wrote and revised the manuscript; Ting Zheng conducted most of the analysis of data; Xuewei Gu reviewed the manuscript; All authors read and approved the final manuscript.

Availability of data and material

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Informed consent was not required from patients to obtain data from the US SEER database since cancer is publicly reportable in every state in the United States.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;70(1):7-30. doi:10.3322/caac.21590

2. Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. 68, 394–424. https ://doi. org/10.3322/caac.21492 (2018).

3. Liu Z, Suo C, Mao X, et al. Global incidence trends in primary liver cancer by age at diagnosis, sex, region, and etiology, 1990-2017. Cancer. 2020;126(10):2267-2278. doi:10.1002 /cncr.32789

4. Kanwal F, Kramer JR, Mapakshi S, et al. Risk of hepatocellular cancer in patients with non-alcoholic fatty liver disease. Gastroenterology. 2018;155:1828-1837.e2.

5. Zhang S, Zheng R, Zeng H, Chen W. The incidence differences among sex, geographical areas and mean age of diagnosis for liver cancer in China, 1989-2008 [in Chinese]. Zhonghua Yu Fang Yi Xue Za Zhi. 2014;48:355-360.

6. Baumeister SE, Schlesinger S, Aleksandrova K, et al. Association between physical activity and risk of hepatobiliary cancers: a multinational cohort study. J Hepatol. 2019;70:885-892.

7. Klil-Drori AJ, Azoulay L, Pollak MN. Cancer, obesity, diabetes, and antidiabetic drugs: is the fog clearing? Nat Rev Clin Oncol. 2017;14:85-99.

8. Smittenaar CR, Petersen KA, Stewart K, Moitt N. Cancer incidence and mortality projections in the UK until 2035. Br J Cancer. 2016;115(9):1147-1155. doi:10.1038/bjc.2016.304

9. Chen, S., et al., Development and Validation of a Nomogram for Predicting Survival in Male Patients With Breast Cancer. Front Oncol, 2019. 9: p. 361

10. Zhang Y, Hong YK, Zhuang DW, He XJ, Lin ME. Bladder cancer survival nomogram: Development and validation of a prediction tool, using the SEER and TCGA databases. Medicine (Baltimore). 2019 Nov;98(44):e17725.

11. Massarweh NN, El-Serag HB. Epidemiology of Hepatocellular Carcinoma and Intrahepatic Cholangiocarcinoma. Cancer Control. 2017;24(3):1073274817729245. doi:10.1177/1073274817729245

12. Fei FR, Hu RY, Gong WW, Pan J, Wang M. Analysis of Mortality and Survival Rate of Liver Cancer in Zhejiang Province in China: A General Population-Based Study. Can J Gastroenterol Hepatol. 2019;2019:1074286. Published 2019 Jul 10. doi:10.1155/2019/1074286

13. Kim SY, Song HK, Lee SK, et al. Sex-Biased Molecular Signature for Overall Survival of Liver Cancer Patients. Biomol Ther (Seoul). 2020;28(6):491-502. doi:10.4062/biomolther.2020.157

14. Gabriele L., Buoncervello M., Ascione B., Bellenghi M., Matarrese P., Carè A. The gender perspective in cancer research and therapy: novel insights and on-going hypotheses. Ann. Ist. Super. Sanita. 2016;52:213–222. doi: 10.4415/ANN_16_02_13.

15. Liu K, Huang G, Chang P, et al. Construction and validation of a nomogram for predicting cancer-specific survival in hepatocellular carcinoma patients. Sci Rep. 2020;10(1):21376. Published 2020 Dec 7. doi:10.1038/s41598-020-78545-2

16. Zhang F, Lu S, Tian M, et al. Albumin-to-Alkaline Phosphatase Ratio is an Independent Prognostic Indicator in Combined Hepatocellular and Cholangiocarcinoma. J Cancer. 2020;11(17):5177-5186. Published 2020 Jun 29. doi:10.7150/jca.45633

17. Ramai D, Ofosu A, Lai JK, Reddy M, Adler DG. Combined Hepatocellular Cholangiocarcinoma: A Population-Based Retrospective Study. Am J Gastroenterol. 2019;114(9):1496-1501. doi:10.14309/ajg.0000000000000326

18. Stavraka C, Rush H, Ross P. Combined hepatocellular cholan-giocarcinoma (c HCC-CC): an update of genetics, molecular biol-ogy, and therapeutic interventions. J Hepatocell Carcinoma,2019,6:11-21.

19. Leoni S, Sansone V, Lorenzo S, et al. Treatment of Combined Hepatocellular and Cholangiocarcinoma. Cancers (Basel). 2020;12(4):794. Published 2020 Mar 26. doi:10.3390/ cancers12040794

20. EW, S. et al. Prediction of indolent prostate cancer: validation and updating of a prognostic nomogram. J. Urol. 177, 107–112; discussion 112, doi:https://doi.org/10.1016/j.jur o.2006.08.068 (2007).

21. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol. 2008;26(8):1364-1370. doi:10.1200/JCO.2007.12.9791

22. Yang J, Li YJ, Liu QQ, Li L, Feng AZ, Wang TY, Zheng S, Xu AD, Lyu J. Brief introduction of medical database and data mining technology in big data era. Journal of Evidence-Based Medicine 2020, 13: 57-69.

Figure Legends

Figure 1 Constructed nomogram

Figure 2 3- and 5-year survival ROC curves for the training and the validation cohorts. A: 3-year survival ROC curve for the training cohort. B: 5-year survival ROC curve for the training cohort. C: 3-year survival ROC curve for the validation cohort. D: 5-year survival ROC curve for the validation cohort.

Figure 3 3- and 5-year survival calibration curves for the training and the validation cohorts. A: 3-year survival calibration curve for the training cohort. B: 5-year survival calibration curve for the training cohort. C: 3-year survival calibration curve for the validation cohort. D: 5-year survival calibration curve for the validation cohort.

Figure 4 The comparison of ROC between the nomogram model and the TNM model. (A: 3-year nomogram model，B: 5-year nomogram model)

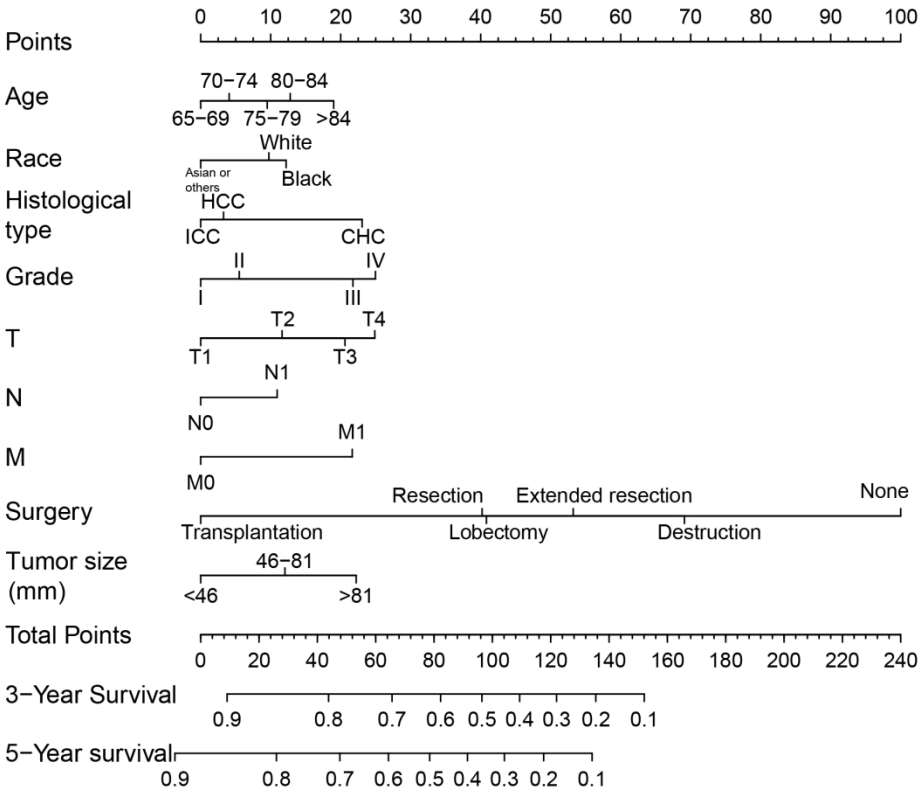


Figure 1 Constructed nomogram

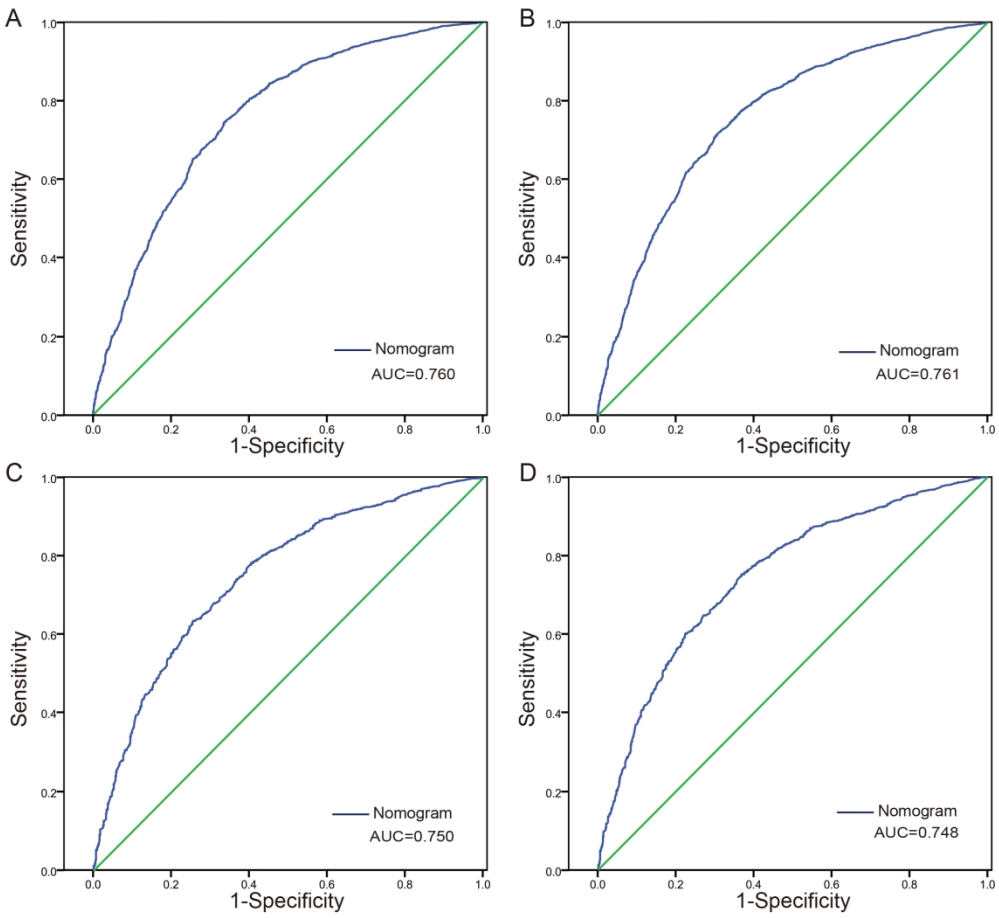


Figure 2 3- and 5-year survival ROC curves for the training and validation cohorts. A: 3-year survival ROC curve for the training cohort. B: 5-year survival ROC curve for the training cohort. C: 3-year survival ROC curve for the validation cohort. D: 5-year survival ROC curve for the validation cohort.

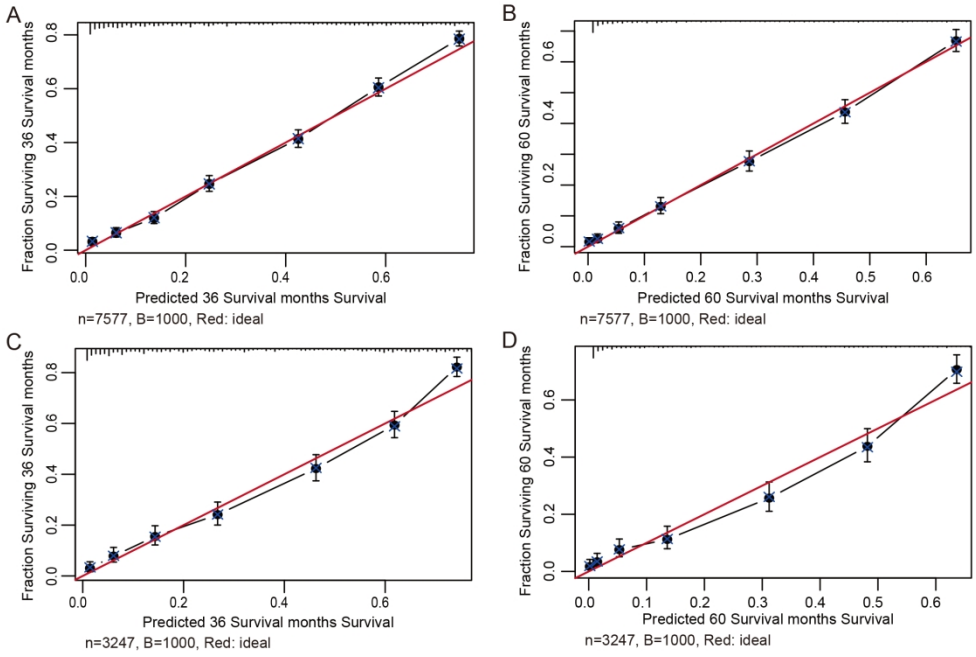


Figure 3 3- and 5-year survival calibration curves for the training and validation cohorts. A: 3-year survival calibration curve for the training cohort. B: 5-year survival calibration curve for the training cohort. C: 3-year survival calibration curve for the validation cohort. D: 5-year survival calibration curve for the validation cohort.

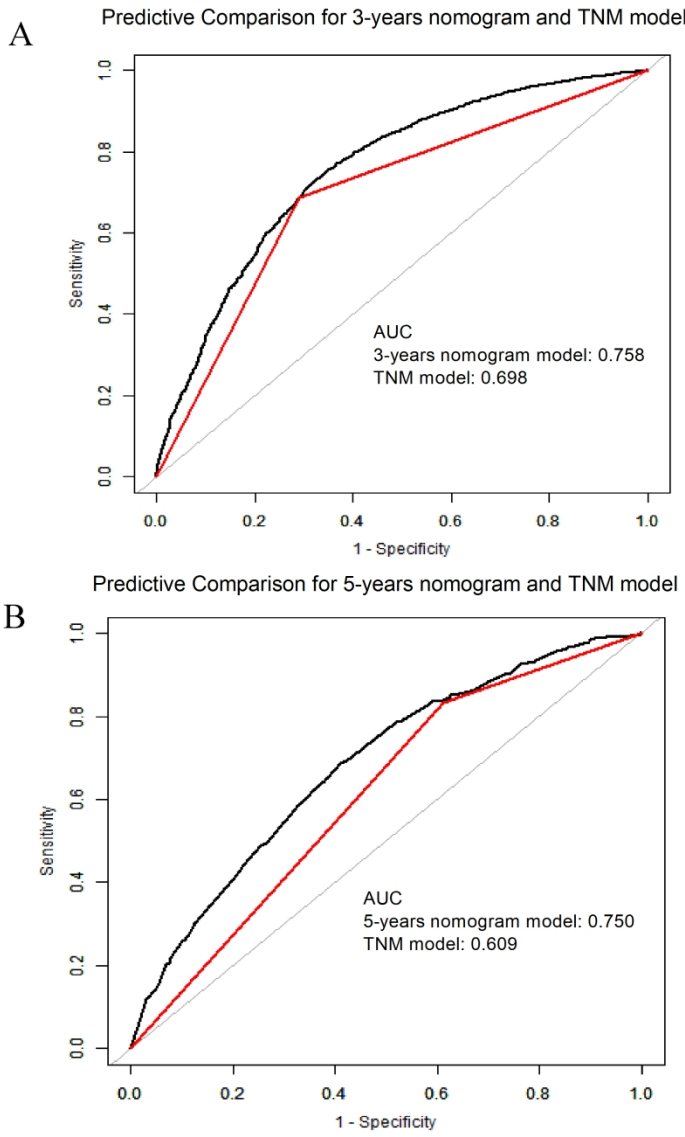


Figure.4 The comparison of ROC between nomogram model and TNM model. (A: 3-years nomogram model, B: 5-years nomogram model)

203x291mm (300 x 300 DPI)

STROBE 2007 (v4) checklist of items to be included in reports of observational studies in epidemiology*
Checklist for cohort, case-control, and cross-sectional studies (combined)

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any pre-specified hypotheses	3
Methods			
Study design	4	Present key elements of study design early in the paper	4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	4
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*4*
Bias	9	Describe any efforts to address potential sources of bias	4
Study size	10	Explain how the study size was arrived at	4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	4
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	5
		(*b*) Describe any methods used to examine subgroups and interactions	5
		(*c*) Explain how missing data were addressed	5
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed	5

		Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	
		(*e*) Describe any sensitivity analyses	
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	5
		(b) Give reasons for non-participation at each stage	
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	6
		(b) Indicate number of participants with missing data for each variable of interest	
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	6
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	6
		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	
Discussion			
Key results	18	Summarise key results with reference to study objectives	7
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	7
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	7
Generalisability	21	Discuss the generalisability (external validity) of the study results	7
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	8

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.
Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.